

**DIACYLHYDRAZINE LIGANDS FOR MODULATING THE EXPRESSION OF
EXOGENOUS GENES IN MAMMALIAN SYSTEMS VIA AN ECDYSONE RECEPTOR
COMPLEX**

This application claims priority to U.S. provisional application No. 60/446,233 filed 2/10/2003.

FIELD OF THE INVENTION

[0001] This invention relates to the field of biotechnology or genetic engineering. Specifically, this invention relates to the field of gene expression. More specifically, this invention relates to non-steroidal ligands for natural and mutated nuclear receptors and their use in a nuclear receptor-based inducible gene expression system and methods of modulating the expression of a gene within a host cell using these ligands and inducible gene expression system.

BACKGROUND OF THE INVENTION

[0002] Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties. However, the citation of any reference herein should not be construed as an admission that such reference is available as "Prior Art" to the instant application.

[0003] In the field of genetic engineering, precise control of gene expression is a valuable tool for studying, manipulating, and controlling development and other physiological processes. Gene expression is a complex biological process involving a number of specific protein-protein interactions. In order for gene expression to be triggered, such that it produces the RNA necessary as the first step in protein synthesis, a transcriptional activator must be brought into proximity of a promoter that controls gene transcription. Typically, the transcriptional activator itself is associated with a protein that has at least one DNA binding domain that binds to DNA binding sites present in the promoter regions of genes. Thus, for gene expression to occur, a protein comprising a DNA binding domain and a transactivation domain located at an appropriate distance from the DNA binding domain must be brought into the correct position in the promoter region of the gene.

[0004] The traditional transgenic approach utilizes a cell-type specific promoter to drive the expression of a designed transgene. A DNA construct containing the transgene is first incorporated into a host genome. When triggered by a transcriptional activator, expression of the transgene occurs in a given cell type.

[0005] Another means to regulate expression of foreign genes in cells is through inducible promoters. Examples of the use of such inducible promoters include the PR1-a promoter, prokaryotic repressor-operator systems, immunosuppressive-immunophilin systems, and higher eukaryotic transcription activation systems such as steroid hormone receptor systems and are described below.

[0006] The PR1-a promoter from tobacco is induced during the systemic acquired resistance response following pathogen attack. The use of PR1-a may be limited because it often responds to endogenous materials and external factors such as pathogens, UV-B radiation, and pollutants. Gene regulation systems based on promoters induced by heat shock, interferon and heavy metals have been described (Wurn et al., 1986, *Proc. Natl. Acad. Sci. USA* 83:5414-5418; Arnheiter et al., 1990 *Cell* 62:51-61; Filmus et al., 1992 *Nucleic Acids Research* 20:27550-27560). However, these systems have limitations due to their effect on expression of non-target genes. These systems are also leaky.

[0007] Prokaryotic repressor-operator systems utilize bacterial repressor proteins and the unique operator DNA sequences to which they bind. Both the tetracycline ("Tet") and lactose ("Lac") repressor-operator systems from the bacterium *Escherichia coli* have been used in plants and animals to control gene expression. In the Tet system, tetracycline binds to the TetR repressor protein, resulting in a conformational change that releases the repressor protein from the operator which as a result allows transcription to occur. In the Lac system, a lac operon is activated in response to the presence of lactose, or synthetic analogs such as isopropyl-b-D-thiogalactoside. Unfortunately, the use of such systems is restricted by unstable chemistry of the ligands, *i.e.* tetracycline and lactose, their toxicity, their natural presence, or the relatively high levels required for induction or repression. For similar reasons, utility of such systems in animals is limited.

[0008] Immunosuppressive molecules such as FK506, rapamycin and cyclosporine A can bind to immunophilins FKBP12, cyclophilin, *etc.* Using this information, a general strategy has been devised to bring together any two proteins simply by placing FK506 on each of the two proteins or by placing FK506 on one and cyclosporine A on another one. A synthetic homodimer of FK506 (FK1012) or a compound resulted from fusion of FK506-cyclosporine (FKCsA) can then be used to induce dimerization of these molecules (Spencer et al., 1993, *Science* 262:1019-24; Belshaw et al., 1996 *Proc Natl Acad Sci U S A* 93:4604-7). Gal4 DNA binding domain fused to FKBP12 and VP16 activator domain fused to cyclophilin, and FKCsA compound were used to show heterodimerization and activation of a reporter gene under the control of a promoter containing Gal4 binding sites. Unfortunately, this system includes immunosuppressants that can have unwanted side effects and therefore, limits its use for various mammalian gene switch applications.

[0009] Higher eukaryotic transcription activation systems such as steroid hormone receptor systems have also been employed. Steroid hormone receptors are members of the nuclear receptor superfamily and are found in vertebrate and invertebrate cells. Unfortunately, use of steroidal compounds that activate the receptors for the regulation of gene expression, particularly in plants and mammals, is limited due to their involvement in many other natural biological pathways in such organisms. In order to overcome such difficulties, an alternative system has been developed using insect ecdysone receptors (EcR).

[0010] Growth, molting, and development in insects are regulated by the ecdysone steroid hormone (molting hormone) and the juvenile hormones (Dhadialla, et al., 1998. Annu. Rev. Entomol. 43: 545-569). The molecular target for ecdysone in insects consists of at least ecdysone receptor (EcR) and ultraspiracle protein (USP). EcR is a member of the nuclear steroid receptor super family that is characterized by signature DNA and ligand binding domains, and an activation domain (Koelle et al. 1991, Cell, 67:59-77). EcR receptors are responsive to a number of steroidal compounds such as ponasterone A and muristerone A. Recently, non-steroidal compounds with ecdysteroid agonist activity have been described, including the commercially available insecticides tebufenozide and methoxyfenozide that are marketed world wide by Rohm and Haas Company (see International Patent Application No. PCT/EP96/00686 and US Patent 5,530,028). Both analogs have exceptional safety profiles to other organisms.

[0011] The insect ecdysone receptor (EcR) heterodimerizes with Ultraspiracle (USP), the insect homologue of the mammalian RXR, and binds ecdysteroids and ecdysone receptor response elements and activate transcription of ecdysone responsive genes. The EcR/USP/ligand complexes play important roles during insect development and reproduction. The EcR is a member of the steroid hormone receptor superfamily and has five modular domains, A/B (transactivation), C (DNA binding, heterodimerization)), D (Hinge, heterodimerization), E (ligand binding, heterodimerization and transactivation and F (transactivation) domains. Some of these domains such as A/B, C and E retain their function when they are fused to other proteins.

[0012] Tightly regulated inducible gene expression systems or "gene switches" are useful for various applications such as gene therapy, large scale production of proteins in cells, cell based high throughput screening assays, functional genomics and regulation of traits in transgenic plants and animals.

[0013] The first version of EcR-based gene switch used *Drosophila melanogaster* EcR (DmEcR) and *Mus musculus* RXR (MmRXR) and showed that these receptors in the presence of steroid, ponasteroneA, transactivate reporter genes in mammalian cell lines and transgenic mice (Christopherson K. S., Mark M.R., Baja J. V., Godowski P. J. 1992, Proc. Natl. Acad. Sci. U.S.A. 89: 6314-6318; No D., Yao T.P., Evans R. M., 1996, Proc. Natl. Acad. Sci. U.S.A. 93: 3346-3351). Later, Suhr et al. 1998, Proc. Natl. Acad. Sci. 95:7999-8004 showed that non-steroidal ecdysone agonist, tebufenozide, induced high level of transactivation of reporter genes in mammalian cells through *Bombyx mori* EcR (BmEcR) in the absence of exogenous heterodimer partner.

[0014] International Patent Applications No. PCT/US97/05330 (WO 97/38117) and PCT/US99/08381 (WO99/58155) disclose methods for modulating the expression of an exogenous gene in which a DNA construct comprising the exogenous gene and an ecdysone response element is activated by a second DNA construct comprising an ecdysone receptor that, in the presence of a ligand therefor, and optionally in the presence of a receptor capable of acting as a silent partner, binds

to the ecdysone response element to induce gene expression. The ecdysone receptor of choice was isolated from *Drosophila melanogaster*. Typically, such systems require the presence of the silent partner, preferably retinoid X receptor (RXR), in order to provide optimum activation. In mammalian cells, insect ecdysone receptor (EcR) heterodimerizes with retinoid X receptor (RXR) and regulates expression of target genes in a ligand dependent manner. International Patent Application No. PCT/US98/14215 (WO 99/02683) discloses that the ecdysone receptor isolated from the silk moth *Bombyx mori* is functional in mammalian systems without the need for an exogenous dimer partner.

[0015] U.S. Patent No. 6,265,173 B1 discloses that various members of the steroid/thyroid superfamily of receptors can combine with *Drosophila melanogaster* ultraspiracle receptor (USP) or fragments thereof comprising at least the dimerization domain of USP for use in a gene expression system. U.S. Patent No. 5,880,333 discloses a *Drosophila melanogaster* EcR and ultraspiracle (USP) heterodimer system used in plants in which the transactivation domain and the DNA binding domain are positioned on two different hybrid proteins. Unfortunately, these USP-based systems are constitutive in animal cells and therefore, are not effective for regulating reporter gene expression.

[0016] In each of these cases, the transactivation domain and the DNA binding domain (either as native EcR as in International Patent Application No. PCT/US98/14215 or as modified EcR as in International Patent Application No. PCT/US97/05330) were incorporated into a single molecule and the other heterodimeric partners, either USP or RXR, were used in their native state.

[0017] Drawbacks of the above described EcR-based gene regulation systems include a considerable background activity in the absence of ligands and non-applicability of these systems for use in both plants and animals (see U.S. Patent No. 5,880,333). Therefore, a need exists in the art for improved EcR-based systems to precisely modulate the expression of exogenous genes in both plants and animals. Such improved systems would be useful for applications such as gene therapy, large-scale production of proteins and antibodies, cell-based high throughput screening assays, functional genomics and regulation of traits in transgenic animals. For certain applications such as gene therapy, it may be desirable to have an inducible gene expression system that responds well to synthetic non-steroid ligands and at the same is insensitive to the natural steroids. Thus, improved systems that are simple, compact, and dependent on ligands that are relatively inexpensive, readily available, and of low toxicity to the host would prove useful for regulating biological systems.

[0018] Recently, it has been shown that an ecdysone receptor-based inducible gene expression system in which the transactivation and DNA binding domains are separated from each other by placing them on two different proteins results in greatly reduced background activity in the absence of a ligand and significantly increased activity over background in the presence of a ligand (pending application PCT/US01/09050, incorporated herein in its entirety by reference). This two-hybrid system is a significantly improved inducible gene expression modulation system compared to the two systems disclosed in applications PCT/US97/05330 and PCT/US98/14215. The two-hybrid system

exploits the ability of a pair of interacting proteins to bring the transcription activation domain into a more favorable position relative to the DNA binding domain such that when the DNA binding domain binds to the DNA binding site on the gene, the transactivation domain more effectively activates the promoter (see, for example, U.S. Patent No. 5,283,173). Briefly, the two-hybrid gene expression system comprises two gene expression cassettes; the first encoding a DNA binding domain fused to a nuclear receptor polypeptide, and the second encoding a transactivation domain fused to a different nuclear receptor polypeptide. In the presence of ligand, the interaction of the first polypeptide with the second polypeptide effectively tethers the DNA binding domain to the transactivation domain. Since the DNA binding and transactivation domains reside on two different molecules, the background activity in the absence of ligand is greatly reduced.

[0019] A two-hybrid system also provides improved sensitivity to non-steroidal ligands for example, diacylhydrazines, when compared to steroidal ligands for example, ponasterone A ("PonA") or muristerone A ("MurA"). That is, when compared to steroids, the non-steroidal ligands provide higher activity at a lower concentration. In addition, since transactivation based on EcR gene switches is often cell-line dependent, it is easier to tailor switching systems to obtain maximum transactivation capability for each application. Furthermore, the two-hybrid system avoids some side effects due to overexpression of RXR that often occur when unmodified RXR is used as a switching partner. In a preferred two-hybrid system, native DNA binding and transactivation domains of EcR or RXR are eliminated and as a result, these hybrid molecules have less chance of interacting with other steroid hormone receptors present in the cell resulting in reduced side effects.

[0020] With the improvement in ecdysone receptor-based gene regulation systems there is an increase in their use in various applications resulting in increased demand for ligands with higher activity than those currently exist. US patent 6,258,603 B1 (and patents cited therein) disclosed dibenzoylhydrazine ligands, however, a need exists for additional ligands with different structures and physicochemical properties. We have discovered novel diacylhydrazine ligands which have not previously been described or shown to have the ability to modulate the expression of transgenes.

SUMMARY OF THE INVENTION

[0021] The present invention relates to non-steroidal ligands for use in nuclear receptor-based inducible gene expression system, and methods of modulating the expression of a gene within a host cell using these ligands with nuclear receptor-based inducible gene expression systems.

[0022] Applicants' invention also relates to methods of modulating gene expression in a host cell using a gene expression modulation system with a ligand of the present invention. Specifically, Applicants' invention provides a method of modulating the expression of a gene in a host cell comprising the steps of: a) introducing into the host cell a gene expression modulation system

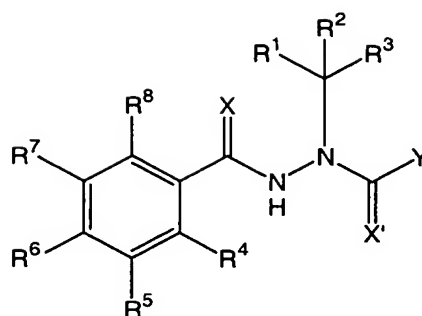
according to the invention; b) introducing into the host cell a gene expression cassette comprising i) a response element comprising a domain to which the DNA binding domain from the first hybrid polypeptide of the gene expression modulation system binds; ii) a promoter that is activated by the transactivation domain of the second hybrid polypeptide of the gene expression modulation system; and iii) a gene whose expression is to be modulated; and c) introducing into the host cell a ligand; whereby upon introduction of the ligand into the host cell, expression of the gene is modulated. Applicants' invention also provides a method of modulating the expression of a gene in a host cell comprising a gene expression cassette comprising a response element comprising a domain to which the DNA binding domain from the first hybrid polypeptide of the gene expression modulation system binds; a promoter that is activated by the transactivation domain of the second hybrid polypeptide of the gene expression modulation system; and a gene whose expression is to be modulated; wherein the method comprises the steps of: a) introducing into the host cell a gene expression modulation system according to the invention; and b) introducing into the host cell a ligand; whereby upon introduction of the ligand into the host, expression of the gene is modulated.

DETAILED DESCRIPTION OF THE INVENTION

[0023] Applicants have discovered novel ligands for natural and mutated nuclear receptors. Thus, Applicants' invention provides a ligand for use with ecdysone receptor-based inducible gene expression system useful for modulating expression of a gene of interest in a host cell. In a particularly desirable embodiment, Applicants' invention provides an inducible gene expression system that has a reduced level of background gene expression and responds to submicromolar concentrations of non-steroidal ligand. Thus, Applicants' novel ligands and inducible gene expression system and its use in methods of modulating gene expression in a host cell overcome the limitations of currently available inducible expression systems and provide the skilled artisan with an effective means to control gene expression.

[0024] The present invention is useful for applications such as gene therapy, large scale production of proteins and antibodies, cell-based high throughput screening assays, functional genomics, proteomics, metabolomics, and regulation of traits in transgenic organisms, where control of gene expression levels is desirable. An advantage of Applicants' invention is that it provides a means to regulate gene expression and to tailor expression levels to suit the user's requirements.

[0025] The present invention pertains to compounds of the general formula:



[0026] wherein X and X' are independently O or S;

[0027] Y is:

- (a) substituted or unsubstituted phenyl wherein the substituents are independently 1-5 H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, (C₂-C₄)alkenyl, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, hydroxy, amino, cyano, or nitro; or
- (b) substituted or unsubstituted 2-pyridyl, 3-pyridyl, or 4-pyridyl, wherein the substituents are independently 1-4 H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, (C₂-C₄)alkenyl, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, hydroxy, amino, cyano, or nitro;

[0028] R¹ and R² are independently: H; cyano; cyano-substituted or unsubstituted (C₁-C₇) branched or straight-chain alkyl; cyano-substituted or unsubstituted (C₂-C₇) branched or straight-chain alkenyl; cyano-substituted or unsubstituted (C₃-C₇) branched or straight-chain alkenylalkyl; or together the valences of R¹ and R² form a (C₁-C₇) cyano-substituted or unsubstituted alkylidene group (R^aR^bC=) wherein the sum of non-substituent carbons in R^a and R^b is 0-6;

[0029] R³ is H, methyl, ethyl, n-propyl, isopropyl, or cyano;

[0030] R⁴, R⁷, and R⁸ are independently: H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, (C₂-C₄)alkenyl, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, hydroxy, amino, cyano, or nitro; and

[0031] R⁵ and R⁶ are independently: H, (C₁-C₄)alkyl, (C₂-C₄)alkenyl, (C₃-C₄) alkenylalkyl, halo (F, Cl, Br, I), C₁-C₄ haloalkyl, (C₁-C₄)alkoxy, hydroxy, amino, cyano, nitro, or together as a linkage of the type (-OCHR⁹CHR¹⁰O-) form a ring with the phenyl carbons to which they are attached; wherein R⁹ and R¹⁰ are independently: H, halo, (C₁-C₃)alkyl, (C₂-C₃)alkenyl, (C₁-C₃)alkoxy(C₁-C₃)alkyl, benzoyloxy(C₁-C₃)alkyl, hydroxy(C₁-C₃)alkyl, halo(C₁-C₃)alkyl, formyl, formyl(C₁-C₃)alkyl, cyano, cyano(C₁-C₃)alkyl, carboxy, carboxy(C₁-C₃)alkyl, (C₁-C₃)alkoxycarbonyl(C₁-C₃)alkyl, (C₁-C₃)alkylcarbonyl(C₁-C₃)alkyl, (C₁-C₃)alkanoyloxy(C₁-C₃)alkyl, amino(C₁-C₃)alkyl, (C₁-C₃)alkylamino(C₁-C₃)alkyl (- (CH₂)_nR^cR^d), oximo (-CH=NOH), oximo(C₁-C₃)alkyl, (C₁-C₃)alkoximo

(-C=NOR^d), alkoximo(C₁-C₃)alkyl, (C₁-C₃)carboxamido (-C(O)NR^eR^f), (C₁-C₃)carboxamido(C₁-C₃)alkyl, (C₁-C₃)semicarbazido (-C=NNHC(O)NR^eR^f), semicarbazido(C₁-C₃)alkyl, aminocarbonyloxy (-OC(O)NHR^g), aminocarbonyloxy(C₁-C₃)alkyl, pentafluorophenyloxycarbonyl, pentafluorophenyloxycarbonyl(C₁-C₃)alkyl, p-toluenesulfonyloxy(C₁-C₃)alkyl, arylsulfonyloxy(C₁-C₃)alkyl, (C₁-C₃)thio(C₁-C₃)alkyl, (C₁-C₃)alkylsulfoxido(C₁-C₃)alkyl, (C₁-C₃)alkylsulfonyl(C₁-C₃)alkyl, or (C₁-C₅)trisubstituted-siloxy(C₁-C₃)alkyl $-(\text{CH}_2)_n\text{SiOR}^d\text{R}^e\text{R}^g$; wherein $n=1-3$, R^e, R^f represent straight or branched hydrocarbon chains of the indicated length, R^e, R^f represent H or straight or branched hydrocarbon chains of the indicated length, R^g represents (C₁-C₃)alkyl or aryl optionally substituted with halo or (C₁-C₃)alkyl, and R^d, R^e, R^f, and R^g are independent of one another;

[0032] provided that

- i when R⁹ and R¹⁰ are both H, or
- ii when either R⁹ or R¹⁰ are halo, (C₁-C₃)alkyl, (C₁-C₃)alkoxy(C₁-C₃)alkyl, or benzoyloxy(C₁-C₃)alkyl, or
- iii when R⁵ and R⁶ do not together form a linkage of the type (-OCHR⁹CHR¹⁰O-),

[0033] then the number of carbon atoms, excluding those of cyano substitution, for either or both of groups R¹ or R² is greater than 4, and the number of carbon atoms, excluding those of cyano substitution, for the sum of groups R¹, R², and R³ is 10, 11, or 12.

[0034] Compounds of the general formula are preferred when:

[0035] X and X' are O;

[0036] Y is:

- (a) substituted or unsubstituted phenyl wherein the substituents are independently 1-5 H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, cyano, or nitro; or
- (b) substituted or unsubstituted 2-pyridyl, 3-pyridyl, or 4-pyridyl, wherein the substituents are independently 1-4 H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, cyano, or nitro;

[0037] R¹ and R² are independently: H; cyano; cyano-substituted or unsubstituted (C₁-C₇) branched or straight-chain alkyl; cyano-substituted or unsubstituted (C₂-C₇) branched or straight-chain alkenyl;

cyano-substituted or unsubstituted (C₃-C₇) branched or straight-chain alkenylalkyl; or together the valences of R¹ and R² form a (C₁-C₇) cyano-substituted or unsubstituted alkylidene group (R^aR^bC=) wherein the sum of non-substituent carbons in R^a and R^b is 0-6;

[0038] R³ is H, methyl, ethyl, or cyano;

[0039] R⁴, R⁷, and R⁸ are independently: H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, cyano, or nitro; and

[0040] R⁵ and R⁶ are independently: H, (C₁-C₄)alkyl, halo (F, Cl, Br, I), C₁-C₄ haloalkyl, (C₁-C₄)alkoxy, hydroxy, amino, cyano, nitro, or together as a linkage of the type (-OCHR⁹CHR¹⁰O-) form a ring with the phenyl carbons to which they are attached; wherein R⁹ or R¹⁰ is H, and the alternate R⁹ or R¹⁰ is: H, halo(C₁-C₃)alkyl, formyl, formyl(C₁-C₃)alkyl, cyano, cyano(C₁-C₃)alkyl, carboxy, carboxy(C₁-C₃)alkyl, amino(C₁-C₃)alkyl, (C₁-C₃)alkylamino(C₁-C₃)alkyl (- (CH₂)_nR^c), oximo (-CH=NOH), oximo(C₁-C₃)alkyl, (C₁-C₃)alkoximo (-C=NOR^d), alkoximo(C₁-C₃)alkyl, (C₁-C₃)carboxamido (-C(O)NR^eR^f), (C₁-C₃)carboxamido(C₁-C₃)alkyl, (C₁-C₃)semicarbazido (-C=NNHC(O)NR^eR^f), semicarbazido(C₁-C₃)alkyl, aminocarbonyloxy (-OC(O)NHR^g), aminocarbonyloxy(C₁-C₃)alkyl, pentafluorophenyloxy, pentafluorophenyloxy(C₁-C₃)alkyl, p-toluenesulfonyloxy(C₁-C₃)alkyl, arylsulfonyloxy(C₁-C₃)alkyl, (C₁-C₃)thio(C₁-C₃)alkyl, (C₁-C₃)alkylsulfoxido(C₁-C₃)alkyl, (C₁-C₃)alkylsulfonyl(C₁-C₃)alkyl, or (C₁-C₃)trisubstituted-siloxy(C₁-C₃)alkyl (- (CH₂)_nSiOR^dR^eR^g); wherein n=1-3, R^c and R^d represent straight or branched hydrocarbon chains of the indicated length, R^e, R^f represent H or straight or branched hydrocarbon chains of the indicated length, R^g represents (C₁-C₃)alkyl or aryl optionally substituted with halo or (C₁-C₃)alkyl, and R^c, R^d, R^e, R^f, and R^g are independent of one another;

[0041] provided that

- i when R⁹ and R¹⁰ are both H, or
- ii when R⁵ and R⁶ do not together form a linkage of the type (-OCHR⁹CHR¹⁰O-),

[0042] then the number of carbon atoms, excluding those of cyano substitution, for either or both of groups R¹ or R² is greater than 4, and the number of carbon atoms, excluding those of cyano substitution, for the sum of groups R¹, R², and R³ is 10, 11, or 12.

[0043] Compounds of the general formula are more preferred when:

[0044] X and X' are O;

[0045] Y is:

- (a) substituted or unsubstituted phenyl wherein the substituents are independently 1-5 H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, cyano, or nitro; or
- (b) substituted or unsubstituted 2-pyridyl, 3-pyridyl, or 4-pyridyl, wherein the substituents are independently 1-4 H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, cyano, or nitro;

[0046] R¹ and R² are independently: H; cyano; cyano-substituted or unsubstituted (C₁-C₇) branched or straight-chain alkyl; cyano-substituted or unsubstituted (C₂-C₇) branched or straight-chain alkenyl; cyano-substituted or unsubstituted (C₃-C₇) branched or straight-chain alkenylalkyl; or together the valences of R¹ and R² form a (C₁-C₇) cyano-substituted or unsubstituted alkylidene group (R^aR^bC=) wherein the sum of non-substituent carbons in R^a and R^b is 0-6;

[0047] R³ is H, methyl, ethyl, or cyano;

[0048] R⁴, R⁷, and R⁸ are independently: H, (C₁-C₄)alkyl, (C₁-C₄)alkoxy, halo (F, Cl, Br, I), (C₁-C₄)haloalkyl, cyano, or nitro; and

[0049] R⁵ and R⁶ are independently: H, (C₁-C₄)alkyl, halo (F, Cl, Br, I), C₁-C₄ haloalkyl, (C₁-C₄)alkoxy, hydroxy, amino, cyano, nitro, or together as a linkage of the type (-OCHR⁹CHR¹⁰O-) form a ring with the phenyl carbons to which they are attached; wherein R⁹ or R¹⁰ is H, and the alternate R⁹ or R¹⁰ is: H, halo(C₁-C₃)alkyl, formyl, formyl(C₁-C₃)alkyl, cyano, cyano(C₁-C₃)alkyl, carboxy, carboxy(C₁-C₃)alkyl, amino(C₁-C₃)alkyl, (C₁-C₃)alkylamino(C₁-C₃)alkyl (-CH₂)_nR^cR^e, oximo (-CH=NOH), oximo(C₁-C₃)alkyl, (C₁-C₃)alkoximo (-C=NOR^d), alkoximo(C₁-C₃)alkyl, (C₁-C₃)carboxamido (-C(O)NR^eR^f), (C₁-C₃)carboxamido(C₁-C₃)alkyl, (C₁-C₃)semicarbazido (-C=NNHC(O)NR^eR^f), semicarbazido(C₁-C₃)alkyl, aminocarbonyloxy (-OC(O)NHR^g), aminocarbonyloxy(C₁-C₃)alkyl, pentafluorophenyloxycarbonyl, pentafluorophenyloxycarbonyl(C₁-C₃)alkyl, p-toluenesulfonyloxy(C₁-C₃)alkyl, arylsulfonyloxy(C₁-C₃)alkyl, (C₁-C₃)thio(C₁-C₃)alkyl, (C₁-C₃)alkylsulfoxido(C₁-C₃)alkyl, (C₁-C₃)alkylsulfonyl(C₁-C₃)alkyl, or (C₁-C₅)trisubstituted-siloxy(C₁-C₃)alkyl (-CH₂)_nSiOR^dR^eR^g; wherein n=1-3, R^c and R^d represent straight or branched hydrocarbon chains of the indicated length, R^e, R^f represent H or straight or branched hydrocarbon chains of the indicated length, R^g represents (C₁-C₃)alkyl or aryl optionally substituted with halo or (C₁-C₃)alkyl, and R^c, R^d, R^e, R^f, and R^g are independent of one another;

[0050] provided that

i when R⁹ and R¹⁰ are both H, or

- ii when R⁵ and R⁶ do not together form a linkage of the type (–OCHR⁹CHR¹⁰O–),

[0051] then the number of carbon atoms, excluding those of cyano substitution, for either or both of groups R¹ or R² is greater than 4, and the number of carbon atoms, excluding those of cyano substitution, for the sum of groups R¹, R², and R³ is 10, 11, or 12; and

[0052] when R⁵ and R⁶ together as a linkage of the type (–OCHR⁹CHR¹⁰O–) form a ring with the phenyl carbons to which they are attached, and R⁹ and R¹⁰ are not both H,

[0053] then R¹ and R² are (C₁–C₄) straight or branched alkyl, and R³ is H or methyl.

[0054] Compounds of the general formula are even more preferred when:

[0055] X and X' are O;

[0056] Y is:

- (a) substituted or unsubstituted phenyl wherein the substituents are independently 1-5 H, (C₁–C₄)alkyl, (C₁–C₄)alkoxy, halo (F, Cl, Br, I), (C₁–C₄)haloalkyl; or
- (b) substituted or unsubstituted 3-pyridyl, wherein the substituents are independently 1-4 H, (C₁–C₄)alkyl, (C₁–C₄)alkoxy, halo (F, Cl, Br, I), (C₁–C₄)haloalkyl;

[0057] R¹ and R² are independently: H; cyano; cyano-substituted or unsubstituted (C₁–C₇) branched or straight-chain alkyl; cyano-substituted or unsubstituted (C₂–C₇) branched or straight-chain alkenyl; cyano-substituted or unsubstituted (C₃–C₇) branched or straight-chain alkenylalkyl; or together the valences of R¹ and R² form a (C₁–C₇) cyano-substituted or unsubstituted alkylidene group (R^aR^bC=) wherein the sum of non-substituent carbons in R^a and R^b is 0-3;

[0058] R³ is methyl;

[0059] R⁴, R⁷, and R⁸ are independently selected from: H, (C₁–C₄)alkyl, (C₁–C₄)alkoxy, halo (F, Cl, Br, I), (C₁–C₄)haloalkyl; and

[0060] R⁵ and R⁶ are independently: H, (C₁–C₄)alkyl, halo (F, Cl, Br, I), C₁–C₄ haloalkyl, (C₁–C₄)alkoxy, or together as a linkage of the type (–OCHR⁹CHR¹⁰O–) form a ring with the phenyl carbons to which they are attached; wherein R⁹ or R¹⁰ is H, and the alternate R⁹ or R¹⁰ is: H, halo(C₁–C₂)alkyl, formyl, cyano(C₁–C₂)alkyl, carboxy, amino(C₁–C₂)alkyl, oximo (–CH=NOH), (C₁–C₃)carboxamido (–C(O)NR^cR^f), (C₁–C₂)semicarbazido (–C=NNHC(O)NR^cR^f), aminocarbonyloxy (–OC(O)NHR^g), pentafluorophenylloxycarbonyl, p-toluenesulfonyloxy(C₁–C₃)alkyl, methylthio(C₁–C₂)alkyl, methylsulfoxido(C₁–C₂)alkyl, methylsulfonyl(C₁–C₂)alkyl, or (C₁–C₃)trisubstituted-siloxy(C₁–C₃)alkyl (–(CH₂)_nSiOR^dR^eR^h); wherein n=1-3, R^d represents a straight or branched

hydrocarbon chain of the indicated length, R^c , R^f represent H or straight or branched hydrocarbon chains of the indicated length, R^g represents (C_1-C_3) alkyl or aryl optionally substituted with halo or (C_1-C_3) alkyl, and R^c , R^d , R^e , R^f , and R^g are independent of one another;

[0061] provided that

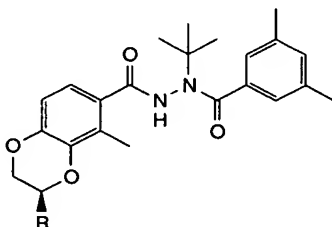
- i) when R^9 and R^{10} are both H, or
- ii) when R^5 and R^6 do not together form a linkage of the type $(-OCHR^9CHR^{10}O-)$,

[0062] then the number of carbon atoms, excluding those of cyano substitution, for either or both of groups R^1 or R^2 is greater than 4, and the number of carbon atoms, excluding those of cyano substitution, for the sum of groups R^1 , R^2 , and R^3 is 10, 11, or 12; and

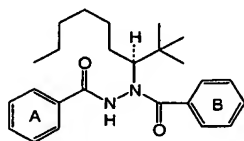
[0063] when R^5 and R^6 together as a linkage of the type $(-OCHR^9CHR^{10}O-)$ form a ring with the phenyl carbons to which they are attached, and R^9 and R^{10} are not both H,

[0064] then R^1 and R^2 are methyl.

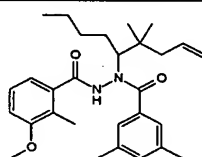
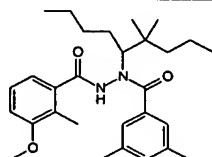
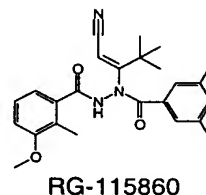
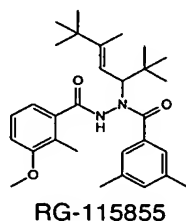
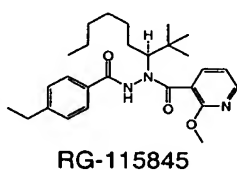
[0065] The compounds of the present invention most preferred are the following:



Compound Reference No.	R
RG-115789	$-CH_2OH$
RG-115790	$-CH_2OSi(tBu)(CH_3)_2$
RG-115805	$-CO_2H$
RG-115806	$-CO_2Me$
RG-115807	$-C=NNHCONH_2$
RG-115808	$-CH_2OC(O)NHPh$
RG-115809	$-CH_2CH_2NH_2$
RG-115810	$-C(O)OC_6F_5$
RG-115811	$-CONHMe$
RG-115812	$-CHO$
RG-115813	$-CH_2OS(O)_2Ph-4-CH_3$
RG-115814	$-C=NOH$
RG-115815	$-CH_2F$
RG-115816	$-CH_2CN$
RG-115817	$-CH_2SCH_3$
RG-115818	$-CH_2S(O)_2CH_3$



Compound Reference No.	A –ring substitution	B –ring substitution
RG-115843	4-Et	3,5-di-CH ₃
RG-115844	4-Et	3,5-di-OCH ₃ , 4-CH ₃
RG-115853	2-CH ₃ , 3-OCH ₃	3,5-di-CH ₃
RG-115854	2-CH ₃ , 3-OCH ₃	3,5-di-OCH ₃ , 4-CH ₃



RG No.	A –ring substitution	B –ring substitution
RG-115845	4-Et	2-OCH ₃ -3-pyridyl
RG-115855	2-CH ₃ , 3-OCH ₃	3,5-di-CH ₃
RG-115860	2-CH ₃ , 3-OCH ₃	3,5-di-CH ₃
RG-115877	2-CH ₃ , 3-OCH ₃	3,5-di-CH ₃
RG-115878	2-CH ₃ , 3-OCH ₃	3,5-di-CH ₃

[0066] Because the compounds of the general formula of the present invention may contain a number of stereogenic carbon atoms, the compounds may exist as enantiomers, diastereomers, stereoisomers, or their mixtures, even if a stereogenic center is explicitly specified.

DEFINITIONS

[0067] When an R^x group is specified, wherein x represents a letter a-g, and the same R^x group is also specified with an alkyl group chain length such as “(C₁-C₃)”, it is understood that the specified chain length refers only to the cases where R^x may be alkyl, and does not pertain to cases where R^x may be a non-alkyl group, such as H or aryl.

[0068] The term "alkyl" includes both branched and straight chain alkyl groups. Typical alkyl groups include, for example, methyl, ethyl, *n*-propyl, isopropyl, *n*-butyl, *sec*-butyl, isobutyl, *tert*-butyl, *n*-pentyl, isopentyl, *n*-hexyl, *n*-heptyl, isooctyl, nonyl, and decyl.

[0069] The term "halo" refers to fluoro, chloro, bromo or iodo.

[0070] The term "haloalkyl" refers to an alkyl group substituted with one or more halo groups such as, for example, chloromethyl, 2-bromoethyl, 3-iodopropyl, trifluoromethyl, and perfluoropropyl.

[0071] The term "cycloalkyl" refers to a cyclic aliphatic ring structure, optionally substituted with alkyl, hydroxy, or halo, such as cyclopropyl, methylcyclopropyl, cyclobutyl, 2-hydroxycyclopentyl, cyclohexyl, and 4-chlorocyclohexyl.

[0072] The term "hydroxyalkyl" refers to an alkyl group substituted with one or more hydroxy groups such as, for example, hydroxymethyl and 2,3-dihydroxybutyl.

[0073] The term "alkylsulfonyl" refers to a sulfonyl moiety substituted with an alkyl group such as, for example, mesyl, and *n*-propylsulfonyl.

[0074] The term "alkenyl" refers to an ethylenically unsaturated hydrocarbon group, straight or branched chain, having 1 or 2 ethylenic bonds such as, for example, vinyl, allyl, 1-butenyl, 2-butenyl, isopropenyl, and 2-pentenyl.

[0075] The term "haloalkenyl" refers to an alkenyl group substituted with one or more halo groups.

[0076] The term "alkynyl" refers to an unsaturated hydrocarbon group, straight or branched, having 1 or 2 acetylenic bonds such as, for example, ethynyl and propargyl.

[0077] The term "alkylcarbonyl" refers to an alkylketo functionality, for example acetyl, *n*-butyryl and the like.

[0078] The term "heterocyclyl" or "heterocycle" refers to an unsubstituted or substituted; saturated, partially unsaturated, or unsaturated 5 or 6-membered ring containing one, two or three heteroatoms, preferably one or two heteroatoms independently selected from the group consisting of oxygen, nitrogen and sulfur. Examples of heterocyclyls include, for example, pyridyl, thienyl, furyl, pyrimidinyl, pyrazinyl, quinolinyl, isoquinolinyl, pyrrolyl, indolyl, tetrahydrofuryl, pyrrolidinyl, piperidinyl, tetrahydropyranyl, morpholinyl, piperazinyl, dioxolanyl, and dioxanyl.

[0079] The term "alkoxy" includes both branched and straight chain alkyl groups attached to a terminal oxygen atom. Typical alkoxy groups include, for example, methoxy, ethoxy, *n*-propoxy, isopropoxy, and *tert*-butoxy.

[0080] The term "haloalkoxy" refers to an alkoxy group substituted with one or more halo groups such as, for example chloromethoxy, trifluoromethoxy, difluoromethoxy, and perfluoroisobutoxy.

[0081] The term "alkylthio" includes both branched and straight chain alkyl groups attached to a terminal sulfur atom such as, for example methylthio.

[0082] The term "haloalkylthio" refers to an alkylthio group substituted with one or more halo groups such as, for example trifluoromethylthio.

[0083] The term "alkoxyalkyl" refers to an alkyl group substituted with an alkoxy group such as, for example, isopropoxymethyl.

[0084] "Silica gel chromatography" refers to a purification method wherein a chemical substance of interest is applied as a concentrated sample to the top of a vertical column of silica gel or chemically-modified silica gel contained in a glass, plastic, or metal cylinder, and elution from such column with a solvent or mixture of solvents.

[0085] "Flash chromatography" refers to silica gel chromatography performed under air, argon, or nitrogen pressure typically in the range of 10 to 50 psi.

[0086] "Gradient chromatography" refers to silica gel chromatography in which the chemical substance is eluted from a column with a progressively changing composition of a solvent mixture.

[0087] "R_F" is a thin layer chromatography term which refers to the fractional distance of movement of a chemical substance of interest on a thin layer chromatography plate, relative to the distance of movement of the eluting solvent system.

[0088] "Parr hydrogenator" and "Parr shaker" refer to apparatus available from Parr Instrument Company, Moline IL, which are designed to facilitate vigorous mixing of a solution containing a chemical substance of interest with an optional solid suspended catalyst and a pressurized, contained atmosphere of a reactant gas. Typically, the gas is hydrogen and the catalyst is palladium, platinum, or oxides thereof deposited on small charcoal particles. The hydrogen pressure is typically in the range of 30 to 70 psi.

[0089] "Dess-Martin reagent" refers to (1,1,1-triacetoxy)-1,1-dihydro-1,2-benziodoxol-3(1H)-one as a solution in dichloromethane available from Acros Organics/Fisher Scientific Company, L.L.C.

[0090] "PS-NMM" refers to a -SO₂NH(CH₂)₃-morpholine functionalized polystyrene resin available from Argonaut Technologies, San Carlos, CA.

[0091] "AP-NCO" refers to an isocyanate-functionalized resin available from Argonaut Technologies, San Carlos, CA.

[0092] "AP-trisamine" refers to a polystyrene-CH₂NHCH₂CH₂NH(CH₂CH₂NH₂)₂ resin available from Argonaut Technologies, San Carlos, CA.

[0093] The term "isolated" for the purposes of the present invention designates a biological material (nucleic acid or protein) that has been removed from its original environment (the environment in which it is naturally present). For example, a polynucleotide present in the natural state in a plant or an animal is not isolated, however the same polynucleotide separated from the adjacent nucleic acids in which it is naturally present, is considered "isolated". The term "purified" does not require the

material to be present in a form exhibiting absolute purity, exclusive of the presence of other compounds. It is rather a relative definition.

[0094] A polynucleotide is in the "purified" state after purification of the starting material or of the natural material by at least one order of magnitude, preferably 2 or 3 and preferably 4 or 5 orders of magnitude.

[0095] A "nucleic acid" is a polymeric compound comprised of covalently linked subunits called nucleotides. Nucleic acid includes polyribonucleic acid (RNA) and polydeoxyribonucleic acid (DNA), both of which may be single-stranded or double-stranded. DNA includes but is not limited to cDNA, genomic DNA, plasmids DNA, synthetic DNA, and semi-synthetic DNA. DNA may be linear, circular, or supercoiled.

[0096] A "nucleic acid molecule" refers to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear or circular DNA molecules (e.g., restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation.

[0097] The term "fragment" will be understood to mean a nucleotide sequence of reduced length relative to the reference nucleic acid and comprising, over the common portion, a nucleotide sequence identical to the reference nucleic acid. Such a nucleic acid fragment according to the invention may be, where appropriate, included in a larger polynucleotide of which it is a constituent. Such fragments comprise, or alternatively consist of, oligonucleotides ranging in length from at least 6, 8, 9, 10, 12, 15, 18, 20, 21, 22, 23, 24, 25, 30, 39, 40, 42, 45, 48, 50, 51, 54, 57, 60, 63, 66, 70, 75, 78, 80, 90, 100, 105, 120, 135, 150, 200, 300, 500, 720, 900, 1000 or 1500 consecutive nucleotides of a nucleic acid according to the invention.

[0098] As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

[0099] A "gene" refers to an assembly of nucleotides that encode a polypeptide, and includes cDNA and genomic DNA nucleic acids. "Gene" also refers to a nucleic acid fragment that expresses a specific protein or polypeptide, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and/or coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. A chimeric gene may comprise coding sequences derived from different sources and/or regulatory sequences derived from different sources. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene or "heterologous" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

[00100] "Heterologous" DNA refers to DNA not naturally located in the cell, or in a chromosomal site of the cell. Preferably, the heterologous DNA includes a gene foreign to the cell.

[00101] The term "genome" includes chromosomal as well as mitochondrial, chloroplast and viral DNA or RNA.

[00102] A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength (see Sambrook *et al.*, 1989 *infra*). Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the "stringency" of the hybridization.

[00103] Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. For preliminary screening for homologous nucleic acids, low stringency hybridization conditions, corresponding to a T_m of 55°, can be used, *e.g.*, 5x SSC, 0.1% SDS, 0.25% milk, and no formamide; or 30% formamide, 5x SSC, 0.5% SDS. Moderate stringency hybridization conditions correspond to a higher T_m , *e.g.*, 40% formamide, with 5x or 6x SCC. High stringency hybridization conditions correspond to the highest T_m , *e.g.*, 50% formamide, 5x or 6x SCC.

[00104] Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The term “complementary” is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, the instant invention also includes isolated nucleic acid fragments that are complementary to the complete sequences as disclosed or used herein as well as those substantially similar nucleic acid sequences.

[00105] In a specific embodiment of the invention, polynucleotides are detected by employing hybridization conditions comprising a hybridization step at T_m of 55°C, and utilizing conditions as set forth above. In a preferred embodiment, the T_m is 60°C; in a more preferred embodiment, the T_m is 63°C; in an even more preferred embodiment, the T_m is 65°C.

[00106] Post-hybridization washes also determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 minutes (min), then repeated with 2X SSC, 0.5% SDS at 45°C for 30 minutes, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 minutes. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C. Hybridization requires that the two nucleic acids comprise complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible.

[00107] The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T_m have been derived (see Sambrook *et al.*, *supra*, 9.50-0.51). For hybridization with shorter nucleic acids, *i.e.*, oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook *et al.*, *supra*, 11.7-11.8).

[00108] In a specific embodiment of the invention, polynucleotides are detected by employing hybridization conditions comprising a hybridization step in less than 500 mM salt and at least 37 degrees Celsius, and a washing step in 2XSSPE at at least 63 degrees Celsius. In a preferred embodiment, the hybridization conditions comprise less than 200 mM salt and at least 37 degrees Celsius for the hybridization step. In a more preferred embodiment, the hybridization conditions comprise 2XSSPE and 63 degrees Celsius for both the hybridization and washing steps.

[00109] In one embodiment, the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferable a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

[00110] The term “probe” refers to a single-stranded nucleic acid molecule that can base pair with a complementary single stranded target nucleic acid to form a double-stranded molecule.

[00111] As used herein, the term “oligonucleotide” refers to a nucleic acid, generally of at least 18 nucleotides, that is hybridizable to a genomic DNA molecule, a cDNA molecule, a plasmid DNA or an mRNA molecule. Oligonucleotides can be labeled, *e.g.*, with ³²P-nucleotides or nucleotides to which a label, such as biotin, has been covalently conjugated. A labeled oligonucleotide can be used as a probe to detect the presence of a nucleic acid. Oligonucleotides (one or both of which may be labeled) can be used as PCR primers, either for cloning full length or a fragment of a nucleic acid, or to detect the presence of a nucleic acid. An oligonucleotide can also be used to form a triple helix with a DNA molecule. Generally, oligonucleotides are prepared synthetically, preferably on a nucleic acid synthesizer. Accordingly, oligonucleotides can be prepared with non-naturally occurring phosphoester analog bonds, such as thioester bonds, etc.

[00112] A “primer” is an oligonucleotide that hybridizes to a target nucleic acid sequence to create a double stranded nucleic acid region that can serve as an initiation point for DNA synthesis under suitable conditions. Such primers may be used in a polymerase chain reaction.

[00113] “Polymerase chain reaction” is abbreviated PCR and means an *in vitro* method for enzymatically amplifying specific nucleic acid sequences. PCR involves a repetitive series of temperature cycles with each cycle comprising three stages: denaturation of the template nucleic acid to separate the strands of the target molecule, annealing a single stranded PCR oligonucleotide primer to the template nucleic acid, and extension of the annealed primer(s) by DNA polymerase. PCR provides a means to detect the presence of the target molecule and, under quantitative or semi-quantitative conditions, to determine the relative amount of that target molecule within the starting pool of nucleic acids.

[00114] “Reverse transcription-polymerase chain reaction” is abbreviated RT-PCR and means an *in vitro* method for enzymatically producing a target cDNA molecule or molecules from an RNA molecule or molecules, followed by enzymatic amplification of a specific nucleic acid sequence or sequences within the target cDNA molecule or molecules as described above. RT-PCR also provides a means to detect the presence of the target molecule and, under quantitative or semi-quantitative conditions, to determine the relative amount of that target molecule within the starting pool of nucleic acids.

[00115] A DNA "coding sequence" is a double-stranded DNA sequence that is transcribed and translated into a polypeptide in a cell *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing site, effector binding site and stem-loop structure. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from mRNA, genomic DNA sequences, and even synthetic DNA sequences. If the coding sequence is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

[00116] "Open reading frame" is abbreviated ORF and means a length of nucleic acid sequence, either DNA, cDNA or RNA, that comprises a translation start signal or initiation codon, such as an ATG or AUG, and a termination codon and can be potentially translated into a polypeptide sequence.

[00117] The term "head-to-head" is used herein to describe the orientation of two polynucleotide sequences in relation to each other. Two polynucleotides are positioned in a head-to-head orientation when the 5' end of the coding strand of one polynucleotide is adjacent to the 5' end of the coding strand of the other polynucleotide, whereby the direction of transcription of each polynucleotide proceeds away from the 5' end of the other polynucleotide. The term "head-to-head" may be abbreviated (5')-to-(5') and may also be indicated by the symbols ($\leftarrow \rightarrow$) or ($3' \leftarrow 5' 5' \rightarrow 3'$).

[00118] The term "tail-to-tail" is used herein to describe the orientation of two polynucleotide sequences in relation to each other. Two polynucleotides are positioned in a tail-to-tail orientation when the 3' end of the coding strand of one polynucleotide is adjacent to the 3' end of the coding strand of the other polynucleotide, whereby the direction of transcription of each polynucleotide proceeds toward the other polynucleotide. The term "tail-to-tail" may be abbreviated (3')-to-(3') and may also be indicated by the symbols ($\rightarrow \leftarrow$) or ($5' \rightarrow 3' 3' \leftarrow 5'$).

[00119] The term "head-to-tail" is used herein to describe the orientation of two polynucleotide sequences in relation to each other. Two polynucleotides are positioned in a head-to-tail orientation when the 5' end of the coding strand of one polynucleotide is adjacent to the 3' end of the coding strand of the other polynucleotide, whereby the direction of transcription of each polynucleotide proceeds in the same direction as that of the other polynucleotide. The term "head-to-tail" may be abbreviated (5')-to-(3') and may also be indicated by the symbols ($\rightarrow \rightarrow$) or ($5' \rightarrow 3' 5' \rightarrow 3'$).

[00120] The term "downstream" refers to a nucleotide sequence that is located 3' to reference nucleotide sequence. In particular, downstream nucleotide sequences generally relate to sequences

that follow the starting point of transcription. For example, the translation initiation codon of a gene is located downstream of the start site of transcription.

[00121] The term “upstream” refers to a nucleotide sequence that is located 5’ to reference nucleotide sequence. In particular, upstream nucleotide sequences generally relate to sequences that are located on the 5’ side of a coding sequence or starting point of transcription. For example, most promoters are located upstream of the start site of transcription.

[00122] The terms “restriction endonuclease” and “restriction enzyme” refer to an enzyme that binds and cuts within a specific nucleotide sequence within double stranded DNA.

[00123] “Homologous recombination” refers to the insertion of a foreign DNA sequence into another DNA molecule, e.g., insertion of a vector in a chromosome. Preferably, the vector targets a specific chromosomal site for homologous recombination. For specific homologous recombination, the vector will contain sufficiently long regions of homology to sequences of the chromosome to allow complementary binding and incorporation of the vector into the chromosome. Longer regions of homology, and greater degrees of sequence similarity, may increase the efficiency of homologous recombination.

[00124] Several methods known in the art may be used to propagate a polynucleotide according to the invention. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. As described herein, the expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (e.g., lambda), and plasmid and cosmid DNA vectors, to name but a few.

[00125] A “vector” is any means for the cloning of and/or transfer of a nucleic acid into a host cell. A vector may be a replicon to which another DNA segment may be attached so as to bring about the replication of the attached segment. A “replicon” is any genetic element (e.g., plasmid, phage, cosmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*, i.e., capable of replication under its own control. The term “vector” includes both viral and nonviral means for introducing the nucleic acid into a cell *in vitro*, *ex vivo* or *in vivo*. A large number of vectors known in the art may be used to manipulate nucleic acids, incorporate response elements and promoters into genes, etc. Possible vectors include, for example, plasmids or modified viruses including, for example bacteriophages such as lambda derivatives, or plasmids such as pBR322 or pUC plasmid derivatives, or the Bluescript vector. For example, the insertion of the DNA fragments corresponding to response elements and promoters into a suitable vector can be accomplished by ligating the appropriate DNA fragments into a chosen vector that has complementary cohesive termini. Alternatively, the ends of the DNA molecules may be enzymatically modified or any site may be produced by ligating nucleotide sequences (linkers) into the DNA termini. Such vectors may

be engineered to contain selectable marker genes that provide for the selection of cells that have incorporated the marker into the cellular genome. Such markers allow identification and/or selection of host cells that incorporate and express the proteins encoded by the marker.

[00126] Viral vectors, and particularly retroviral vectors, have been used in a wide variety of gene delivery applications in cells, as well as living animal subjects. Viral vectors that can be used include but are not limited to retrovirus, adeno-associated virus, pox, baculovirus, vaccinia, herpes simplex, Epstein-Barr, adenovirus, geminivirus, and caulimovirus vectors. Non-viral vectors include plasmids, liposomes, electrically charged lipids (cytofectins), DNA-protein complexes, and biopolymers. In addition to a nucleic acid, a vector may also comprise one or more regulatory regions, and/or selectable markers useful in selecting, measuring, and monitoring nucleic acid transfer results (transfer to which tissues, duration of expression, etc.).

[00127] The term "plasmid" refers to an extra chromosomal element often carrying a gene that is not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear, circular, or supercoiled, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell.

[00128] A "cloning vector" is a "replicon", which is a unit length of a nucleic acid, preferably DNA, that replicates sequentially and which comprises an origin of replication, such as a plasmid, phage or cosmid, to which another nucleic acid segment may be attached so as to bring about the replication of the attached segment. Cloning vectors may be capable of replication in one cell type and expression in another ("shuttle vector").

[00129] Vectors may be introduced into the desired host cells by methods known in the art, *e.g.*, transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, lipofection (lysosome fusion), use of a gene gun, or a DNA vector transporter (see, *e.g.*, Wu et al., 1992, J. Biol. Chem. 267: 963-967; Wu and Wu, 1988, J. Biol. Chem. 263: 14621-14624; and Hartmut et al., Canadian Patent Application No. 2,012,311, filed March 15, 1990).

[00130] A polynucleotide according to the invention can also be introduced *in vivo* by lipofection. For the past decade, there has been increasing use of liposomes for encapsulation and transfection of nucleic acids *in vitro*. Synthetic cationic lipids designed to limit the difficulties and dangers encountered with liposome-mediated transfection can be used to prepare liposomes for *in vivo* transfection of a gene encoding a marker (Felgner et al., 1987, PNAS 84:7413; Mackey, et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:8027-8031; and Ulmer et al., 1993, Science 259:1745-1748). The use of cationic lipids may promote encapsulation of negatively charged nucleic acids, and also

promote fusion with negatively charged cell membranes (Felgner and Ringold, 1989, Science 337: 387-388). Particularly useful lipid compounds and compositions for transfer of nucleic acids are described in International Patent Publications WO95/18863 and WO96/17823, and in U.S. Patent No. 5,459,127. The use of lipofection to introduce exogenous genes into the specific organs *in vivo* has certain practical advantages. Molecular targeting of liposomes to specific cells represents one area of benefit. It is clear that directing transfection to particular cell types would be particularly preferred in a tissue with cellular heterogeneity, such as pancreas, liver, kidney, and the brain. Lipids may be chemically coupled to other molecules for the purpose of targeting (Mackey, et al., 1988, *supra*). Targeted peptides, *e.g.*, hormones or neurotransmitters, and proteins such as antibodies, or non-peptide molecules could be coupled to liposomes chemically.

[00131] Other molecules are also useful for facilitating transfection of a nucleic acid *in vivo*, such as a cationic oligopeptide (*e.g.*, WO95/21931), peptides derived from DNA binding proteins (*e.g.*, WO96/25508), or a cationic polymer (*e.g.*, WO95/21931).

[00132] It is also possible to introduce a vector *in vivo* as a naked DNA plasmid (see U.S. Patents 5,693,622, 5,589,466 and 5,580,859). Receptor-mediated DNA delivery approaches can also be used (Curiel et al., 1992, Hum. Gene Ther. 3: 147-154; and Wu and Wu, 1987, J. Biol. Chem. 262: 4429-4432).

[00133] The term "transfection" means the uptake of exogenous or heterologous RNA or DNA by a cell. A cell has been "transfected" by exogenous or heterologous RNA or DNA when such RNA or DNA has been introduced inside the cell. A cell has been "transformed" by exogenous or heterologous RNA or DNA when the transfected RNA or DNA effects a phenotypic change. The transforming RNA or DNA can be integrated (covalently linked) into chromosomal DNA making up the genome of the cell.

[00134] "Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" or "recombinant" or "transformed" organisms.

[00135] The term "genetic region" will refer to a region of a nucleic acid molecule or a nucleotide sequence that comprises a gene encoding a polypeptide.

[00136] In addition, the recombinant vector comprising a polynucleotide according to the invention may include one or more origins for replication in the cellular hosts in which their amplification or their expression is sought, markers or selectable markers.

[00137] The term "selectable marker" means an identifying factor, usually an antibiotic or chemical resistance gene, that is able to be selected for based upon the marker gene's effect, *i.e.*, resistance to an antibiotic, resistance to a herbicide, colorimetric markers, enzymes, fluorescent markers, and the like, wherein the effect is used to track the inheritance of a nucleic acid of interest and/or to identify a cell or organism that has inherited the nucleic acid of interest. Examples of selectable marker genes

known and used in the art include: genes providing resistance to ampicillin, streptomycin, gentamycin, kanamycin, hygromycin, bialaphos herbicide, sulfonamide, and the like; and genes that are used as phenotypic markers, *i.e.*, anthocyanin regulatory genes, isopentanyl transferase gene, and the like.

[00138] The term “reporter gene” means a nucleic acid encoding an identifying factor that is able to be identified based upon the reporter gene’s effect, wherein the effect is used to track the inheritance of a nucleic acid of interest, to identify a cell or organism that has inherited the nucleic acid of interest, and/or to measure gene expression induction or transcription. Examples of reporter genes known and used in the art include: luciferase (Luc), green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), β -galactosidase (LacZ), β -glucuronidase (Gus), and the like. Selectable marker genes may also be considered reporter genes.

[00139] “Promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. Promoters that cause a gene to be expressed in a specific cell type are commonly referred to as “cell-specific promoters” or “tissue-specific promoters”. Promoters that cause a gene to be expressed at a specific stage of development or cell differentiation are commonly referred to as “developmentally-specific promoters” or “cell differentiation-specific promoters”. Promoters that are induced and cause a gene to be expressed following exposure or treatment of the cell with an agent, biological molecule, chemical, ligand, light, or the like that induces the promoter are commonly referred to as “inducible promoters” or “regulatable promoters”. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

[00140] A “promoter sequence” is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined for example, by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

[00141] A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced (if the coding sequence contains introns) and translated into the protein encoded by the coding sequence.

[00142] "Transcriptional and translational control sequences" are DNA regulatory sequences, such as promoters, enhancers, terminators, and the like, that provide for the expression of a coding sequence in a host cell. In eukaryotic cells, polyadenylation signals are control sequences.

[00143] The term "response element" means one or more cis-acting DNA elements which confer responsiveness on a promoter mediated through interaction with the DNA-binding domains of the first chimeric gene. This DNA element may be either palindromic (perfect or imperfect) in its sequence or composed of sequence motifs or half sites separated by a variable number of nucleotides. The half sites can be similar or identical and arranged as either direct or inverted repeats or as a single half site or multimers of adjacent half sites in tandem. The response element may comprise a minimal promoter isolated from different organisms depending upon the nature of the cell or organism into which the response element will be incorporated. The DNA binding domain of the first hybrid protein binds, in the presence or absence of a ligand, to the DNA sequence of a response element to initiate or suppress transcription of downstream gene(s) under the regulation of this response element.

Examples of DNA sequences for response elements of the natural ecdysone receptor include:

RRGG/TTCANTGAC/ACY (see Cherbas L., et. al., (1991), *Genes Dev.* 5, 120-131);

AGGTCAN_(n)AGGTCA, where N_(n) can be one or more spacer nucleotides (see D'Avino PP., et. al., (1995), *Mol. Cell. Endocrinol.* 113, 1-9); and GGGTTGAATGAATTT (see Antoniewski C., et. al., (1994). *Mol. Cell Biol.* 14, 4465-4474).

[00144] The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

[00145] The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from a nucleic acid or polynucleotide. Expression may also refer to translation of mRNA into a protein or polypeptide.

[00146] The terms "cassette", "expression cassette" and "gene expression cassette" refer to a segment of DNA that can be inserted into a nucleic acid or polynucleotide at specific restriction sites or by homologous recombination. The segment of DNA comprises a polynucleotide that encodes a polypeptide of interest, and the cassette and restriction sites are designed to ensure insertion of the cassette in the proper reading frame for transcription and translation. "Transformation cassette" refers to a specific vector comprising a polynucleotide that encodes a polypeptide of interest and having

elements in addition to the polynucleotide that facilitate transformation of a particular host cell. Cassettes, expression cassettes, gene expression cassettes and transformation cassettes of the invention may also comprise elements that allow for enhanced expression of a polynucleotide encoding a polypeptide of interest in a host cell. These elements may include, but are not limited to: a promoter, a minimal promoter, an enhancer, a response element, a terminator sequence, a polyadenylation sequence, and the like.

[00147] For purposes of this invention, the term “gene switch” refers to the combination of a response element associated with a promoter, and an EcR based system which in the presence of one or more ligands, modulates the expression of a gene into which the response element and promoter are incorporated.

[00148] The terms “modulate” and “modulates” mean to induce, reduce or inhibit nucleic acid or gene expression, resulting in the respective induction, reduction or inhibition of protein or polypeptide production.

[00149] The plasmids or vectors according to the invention may further comprise at least one promoter suitable for driving expression of a gene in a host cell. The term “expression vector” means a vector, plasmid or vehicle designed to enable the expression of an inserted nucleic acid sequence following transformation into the host. The cloned gene, i.e., the inserted nucleic acid sequence, is usually placed under the control of control elements such as a promoter, a minimal promoter, an enhancer, or the like. Initiation control regions or promoters, which are useful to drive expression of a nucleic acid in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genes is suitable for the present invention including but not limited to: viral promoters, bacterial promoters, animal promoters, mammalian promoters, synthetic promoters, constitutive promoters, tissue specific promoter, developmental specific promoters, inducible promoters, light regulated promoters; *CYC1*, *HIS3*, *GAL1*, *GAL4*, *GAL10*, *ADHI*, *PGK*, *PHO5*, *GAPDH*, *ADC1*, *TRP1*, *URA3*, *LEU2*, *ENO*, *TPI*, alkaline phosphatase promoters (useful for expression in *Saccharomyces*); *AOX1* promoter (useful for expression in *Pichia*); β -lactamase, *lac*, *ara*, *tet*, *trp*, *lP_L*, *lP_R*, *T7*, *tac*, and *trc* promoters (useful for expression in *Escherichia coli*); light regulated-, seed specific-, pollen specific-, ovary specific-, pathogenesis or disease related-, cauliflower mosaic virus 35S, CMV 35S minimal, cassava vein mosaic virus (CsVMV), chlorophyll a/b binding protein, ribulose 1, 5-bisphosphate carboxylase, shoot-specific, root specific, chitinase, stress inducible, rice tungro bacilliform virus, plant super-promoter, potato leucine aminopeptidase, nitrate reductase, mannopine synthase, nopaline synthase, ubiquitin, zein protein, and anthocyanin promoters (useful for expression in plant cells); animal and mammalian promoters known in the art include, but are not limited to, the SV40 early (SV40e) promoter region, the promoter contained in the 3' long terminal repeat (LTR) of Rous sarcoma virus (RSV), the promoters of the E1A or major late promoter (MLP) genes of adenoviruses (Ad), the cytomegalovirus (CMV) early promoter, the herpes

simplex virus (HSV) thymidine kinase (TK) promoter, a baculovirus IE1 promoter, an elongation factor 1 alpha (EF1) promoter, a phosphoglycerate kinase (PGK) promoter, a ubiquitin (Ubc) promoter, an albumin promoter, the regulatory sequences of the mouse metallothionein-L promoter and transcriptional control regions, the ubiquitous promoters (HPRT, vimentin, α -actin, tubulin and the like), the promoters of the intermediate filaments (desmin, neurofilaments, keratin, GFAP, and the like), the promoters of therapeutic genes (of the MDR, CFTR or factor VIII type, and the like), pathogenesis or disease related-promoters, and promoters that exhibit tissue specificity and have been utilized in transgenic animals, such as the elastase I gene control region which is active in pancreatic acinar cells; insulin gene control region active in pancreatic beta cells, immunoglobulin gene control region active in lymphoid cells, mouse mammary tumor virus control region active in testicular, breast, lymphoid and mast cells; albumin gene, Apo AI and Apo AII control regions active in liver, alpha-fetoprotein gene control region active in liver, alpha 1-antitrypsin gene control region active in the liver, beta-globin gene control region active in myeloid cells, myelin basic protein gene control region active in oligodendrocyte cells in the brain, myosin light chain-2 gene control region active in skeletal muscle, and gonadotropic releasing hormone gene control region active in the hypothalamus, pyruvate kinase promoter, villin promoter, promoter of the fatty acid binding intestinal protein, promoter of the smooth muscle cell α -actin, and the like. In addition, these expression sequences may be modified by addition of enhancer or regulatory sequences and the like.

[00150] Enhancers that may be used in embodiments of the invention include but are not limited to: an SV40 enhancer, a cytomegalovirus (CMV) enhancer, an elongation factor 1 (EF1) enhancer, yeast enhancers, viral gene enhancers, and the like.

[00151] Termination control regions, *i.e.*, terminator or polyadenylation sequences, may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary, however, it is most preferred if included. In a preferred embodiment of the invention, the termination control region may be comprise or be derived from a synthetic sequence, synthetic polyadenylation signal, an SV40 late polyadenylation signal, an SV40 polyadenylation signal, a bovine growth hormone (BGH) polyadenylation signal, viral terminator sequences, or the like.

[00152] The terms “3' non-coding sequences” or “3' untranslated region (UTR)” refer to DNA sequences located downstream (3') of a coding sequence and may comprise polyadenylation [poly(A)] recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

[00153] “Regulatory region” means a nucleic acid sequence that regulates the expression of a second nucleic acid sequence. A regulatory region may include sequences which are naturally responsible for expressing a particular nucleic acid (a homologous region) or may include sequences of a different origin that are responsible for expressing different proteins or even synthetic proteins (a

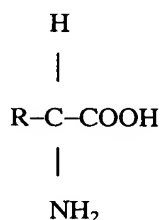
heterologous region). In particular, the sequences can be sequences of prokaryotic, eukaryotic, or viral genes or derived sequences that stimulate or repress transcription of a gene in a specific or non-specific manner and in an inducible or non-inducible manner. Regulatory regions include origins of replication, RNA splice sites, promoters, enhancers, transcriptional termination sequences, and signal sequences which direct the polypeptide into the secretory pathways of the target cell.

[00154] A regulatory region from a “heterologous source” is a regulatory region that is not naturally associated with the expressed nucleic acid. Included among the heterologous regulatory regions are regulatory regions from a different species, regulatory regions from a different gene, hybrid regulatory sequences, and regulatory sequences which do not occur in nature, but which are designed by one having ordinary skill in the art.

[00155] “RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA.

“Messenger RNA (mRNA)” refers to the RNA that is without introns and that can be translated into protein by the cell. “cDNA” refers to a double-stranded DNA that is complementary to and derived from mRNA. “Sense” RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. “Antisense RNA” refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene. The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, or the coding sequence. “Functional RNA” refers to antisense RNA, ribozyme RNA, or other RNA that is not translated yet has an effect on cellular processes.

[00156] A “polypeptide” is a polymeric compound comprised of covalently linked amino acid residues. Amino acids have the following general structure:



[00157] Amino acids are classified into seven groups on the basis of the side chain R: (1) aliphatic side chains, (2) side chains containing a hydroxylic (OH) group, (3) side chains containing sulfur atoms, (4) side chains containing an acidic or amide group, (5) side chains containing a basic group, (6) side chains containing an aromatic ring, and (7) proline, an imino acid in which the side chain is fused to the amino group. A polypeptide of the invention preferably comprises at least about 14 amino acids.

[00158] A “protein” is a polypeptide that performs a structural or functional role in a living cell.

[00159] An “isolated polypeptide” or “isolated protein” is a polypeptide or protein that is substantially free of those compounds that are normally associated therewith in its natural state (e.g., other proteins or polypeptides, nucleic acids, carbohydrates, lipids). “Isolated” is not meant to exclude artificial or synthetic mixtures with other compounds, or the presence of impurities which do not interfere with biological activity, and which may be present, for example, due to incomplete purification, addition of stabilizers, or compounding into a pharmaceutically acceptable preparation.

[00160] A “substitution mutant polypeptide” or a “substitution mutant” will be understood to mean a mutant polypeptide comprising a substitution of at least one (1) wild-type or naturally occurring amino acid with a different amino acid relative to the wild-type or naturally occurring polypeptide. A substitution mutant polypeptide may comprise only one (1) wild-type or naturally occurring amino acid substitution and may be referred to as a “point mutant” or a “single point mutant” polypeptide. Alternatively, a substitution mutant polypeptide may comprise a substitution of two (2) or more wild-type or naturally occurring amino acids with 2 or more amino acids relative to the wild-type or naturally occurring polypeptide. According to the invention, a Group H nuclear receptor ligand binding domain polypeptide comprising a substitution mutation comprises a substitution of at least one (1) wild-type or naturally occurring amino acid with a different amino acid relative to the wild-type or naturally occurring Group H nuclear receptor ligand binding domain polypeptide.

[00161] Wherein the substitution mutant polypeptide comprises a substitution of two (2) or more wild-type or naturally occurring amino acids, this substitution may comprise either an equivalent number of wild-type or naturally occurring amino acids deleted for the substitution, i.e., 2 wild-type or naturally occurring amino acids replaced with 2 non-wild-type or non-naturally occurring amino acids, or a non-equivalent number of wild-type amino acids deleted for the substitution, i.e., 2 wild-type amino acids replaced with 1 non-wild-type amino acid (a substitution+deletion mutation), or 2 wild-type amino acids replaced with 3 non-wild-type amino acids (a substitution+insertion mutation).

[00162] Substitution mutants may be described using an abbreviated nomenclature system to indicate the amino acid residue and number replaced within the reference polypeptide sequence and the new substituted amino acid residue. For example, a substitution mutant in which the twentieth (20th) amino acid residue of a polypeptide is substituted may be abbreviated as “x20z”, wherein “x” is the amino acid to be replaced, “20” is the amino acid residue position or number within the polypeptide, and “z” is the new substituted amino acid. Therefore, a substitution mutant abbreviated interchangeably as “E20A” or “Glu20Ala” indicates that the mutant comprises an alanine residue (commonly abbreviated in the art as “A” or “Ala”) in place of the glutamic acid (commonly abbreviated in the art as “E” or “Glu”) at position 20 of the polypeptide.

[00163] A substitution mutation may be made by any technique for mutagenesis known in the art, including but not limited to, *in vitro* site-directed mutagenesis (Hutchinson, C., et al., 1978, J. Biol.

Chem. 253: 6551; Zoller and Smith, 1984, DNA 3: 479-488; Oliphant et al., 1986, Gene 44: 177; Hutchinson et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83: 710), use of TAB® linkers (Pharmacia), restriction endonuclease digestion/fragment deletion and substitution, PCR-mediated/oligonucleotide-directed mutagenesis, and the like. PCR-based techniques are preferred for site-directed mutagenesis (see Higuchi, 1989, "Using PCR to Engineer DNA", in *PCR Technology: Principles and Applications for DNA Amplification*, H. Erlich, ed., Stockton Press, Chapter 6, pp. 61-70).

[00164] "Fragment" of a polypeptide according to the invention will be understood to mean a polypeptide whose amino acid sequence is shorter than that of the reference polypeptide and which comprises, over the entire portion with these reference polypeptides, an identical amino acid sequence. Such fragments may, where appropriate, be included in a larger polypeptide of which they are a part. Such fragments of a polypeptide according to the invention may have a length of at least 2, 3, 4, 5, 6, 8, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 30, 35, 40, 45, 50, 100, 200, 240, or 300 amino acids.

[00165] A "variant" of a polypeptide or protein is any analogue, fragment, derivative, or mutant which is derived from a polypeptide or protein and which retains at least one biological property of the polypeptide or protein. Different variants of the polypeptide or protein may exist in nature. These variants may be allelic variations characterized by differences in the nucleotide sequences of the structural gene coding for the protein, or may involve differential splicing or post-translational modification. The skilled artisan can produce variants having single or multiple amino acid substitutions, deletions, additions, or replacements. These variants may include, *inter alia*: (a) variants in which one or more amino acid residues are substituted with conservative or non-conservative amino acids, (b) variants in which one or more amino acids are added to the polypeptide or protein, (c) variants in which one or more of the amino acids includes a substituent group, and (d) variants in which the polypeptide or protein is fused with another polypeptide such as serum albumin. The techniques for obtaining these variants, including genetic (suppressions, deletions, mutations, etc.), chemical, and enzymatic techniques, are known to persons having ordinary skill in the art. A variant polypeptide preferably comprises at least about 14 amino acids.

[00166] A "heterologous protein" refers to a protein not naturally produced in the cell.

[00167] A "mature protein" refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed.

"Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

[00168] The term "signal peptide" refers to an amino terminal polypeptide preceding the secreted mature protein. The signal peptide is cleaved from and is therefore not present in the mature protein.

Signal peptides have the function of directing and translocating secreted proteins across cell membranes. Signal peptide is also referred to as signal protein.

[00169] A "signal sequence" is included at the beginning of the coding sequence of a protein to be expressed on the surface of a cell. This sequence encodes a signal peptide, N-terminal to the mature polypeptide, that directs the host cell to translocate the polypeptide. The term "translocation signal sequence" is used herein to refer to this sort of signal sequence. Translocation signal sequences can be found associated with a variety of proteins native to eukaryotes and prokaryotes, and are often functional in both types of organisms.

[00170] The term "homology" refers to the percent of identity between two polynucleotide or two polypeptide moieties. The correspondence between the sequence from one moiety to another can be determined by techniques known to the art. For example, homology can be determined by a direct comparison of the sequence information between two polypeptide molecules by aligning the sequence information and using readily available computer programs. Alternatively, homology can be determined by hybridization of polynucleotides under conditions that form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s) and size determination of the digested fragments.

[00171] As used herein, the term "homologous" in all its grammatical forms and spelling variations refers to the relationship between proteins that possess a "common evolutionary origin," including proteins from superfamilies (*e.g.*, the immunoglobulin superfamily) and homologous proteins from different species (*e.g.*, myosin light chain, etc.) (Reeck et al., 1987, Cell 50:667.). Such proteins (and their encoding genes) have sequence homology, as reflected by their high degree of sequence similarity. However, in common usage and in the instant application, the term "homologous," when modified with an adverb such as "highly," may refer to sequence similarity and not a common evolutionary origin.

[00172] Accordingly, the term "sequence similarity" in all its grammatical forms refers to the degree of identity or correspondence between nucleic acid or amino acid sequences of proteins that may or may not share a common evolutionary origin (*see* Reeck et al., 1987, Cell 50: 667).

[00173] In a specific embodiment, two DNA sequences are "substantially homologous" or "substantially similar" when at least about 50% (preferably at least about 75%, and most preferably at least about 90 or 95%) of the nucleotides match over the defined length of the DNA sequences. Sequences that are substantially homologous can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, *e.g.*, Sambrook *et al.*, 1989, *supra*.

[00174] As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the

functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotide bases that do not substantially affect the functional properties of the resulting transcript. It is therefore understood that the invention encompasses more than the specific exemplary sequences. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products.

[00175] Moreover, the skilled artisan recognizes that substantially similar sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS), with the sequences exemplified herein. Substantially similar nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are at least 70% identical to the DNA sequence of the nucleic acid fragments reported herein. Preferred substantially nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are at least 80% identical to the DNA sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are at least 90% identical to the DNA sequence of the nucleic acid fragments reported herein. Even more preferred are nucleic acid fragments that are at least 95% identical to the DNA sequence of the nucleic acid fragments reported herein.

[00176] Two amino acid sequences are "substantially homologous" or "substantially similar" when greater than about 40% of the amino acids are identical, or greater than 60% are similar (functionally identical). Preferably, the similar or homologous sequences are identified by alignment using, for example, the GCG (Genetics Computer Group, Program Manual for the GCG Package, *Version 7*, Madison, Wisconsin) pileup program.

[00177] The term "corresponding to" is used herein to refer to similar or homologous sequences, whether the exact position is identical or different from the molecule to which the similarity or homology is measured. A nucleic acid or amino acid sequence alignment may include spaces. Thus, the term "corresponding to" refers to the sequence similarity, and not the numbering of the amino acid residues or nucleotide bases.

[00178] A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215: 403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or

thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a “substantial portion” of a nucleotide sequence comprises enough of the sequence to specifically identify and/or isolate a nucleic acid fragment comprising the sequence.

[00179] The term “percent identity”, as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, “identity” also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. “Identity” and “similarity” can be readily calculated by known methods, including but not limited to those described in: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, New York (1993); *Computer Analysis of Sequence Data, Part I* (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press (1987); and *Sequence Analysis Primer* (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences may be performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method may be selected: KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

[00180] The term “sequence analysis software” refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. “Sequence analysis software” may be commercially available or independently developed. Typical sequence analysis software will include but is not limited to the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), and DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 USA). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the “default

values” of the program referenced, unless otherwise specified. As used herein “default values” will mean any set of values or parameters which originally load with the software when first initialized.

[00181] “Synthetic genes” can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments that are then enzymatically assembled to construct the entire gene. “Chemically synthesized”, as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well-established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

[00182] As used herein, two or more individually operable gene regulation systems are said to be “orthogonal” when; a) modulation of each of the given systems by its respective ligand, at a chosen concentration, results in a measurable change in the magnitude of expression of the gene of that system, and b) the change is statistically significantly different than the change in expression of all other systems simultaneously operable in the cell, tissue, or organism, regardless of the simultaneity or sequentially of the actual modulation. Preferably, modulation of each individually operable gene regulation system effects a change in gene expression at least 2-fold greater than all other operable systems in the cell, tissue, or organism. More preferably, the change is at least 5-fold greater. Even more preferably, the change is at least 10-fold greater. Still more preferably, the change is at least 100 fold greater. Even still more preferably, the change is at least 500-fold greater. Ideally, modulation of each of the given systems by its respective ligand at a chosen concentration results in a measurable change in the magnitude of expression of the gene of that system and no measurable change in expression of all other systems operable in the cell, tissue, or organism. In such cases the multiple inducible gene regulation system is said to be “fully orthogonal”. The present invention is useful to search for orthogonal ligands and orthogonal receptor-based gene expression systems such as those described in co-pending US application 09/965,697, which is incorporated herein by reference in its entirety.

[00183] The term “modulate” means the ability of a given ligand/receptor complex to induce or suppress the transactivation of an exogenous gene.

[00184] The term “exogenous gene” means a gene foreign to the subject, that is, a gene which is introduced into the subject through a transformation process, an unmutated version of an endogenous mutated gene or a mutated version of an endogenous unmutated gene. The method of transformation is not critical to this invention and may be any method suitable for the subject known to those in the

art. For example, transgenic plants are obtained by regeneration from the transformed cells. Numerous transformation procedures are known from the literature such as agroinfection using *Agrobacterium tumefaciens* or its T₁ plasmid, electroporation, microinjection of plant cells and protoplasts, and microprojectile transformation. Complementary techniques are known for transformation of animal cells and regeneration of such transformed cells in transgenic animals. Exogenous genes can be either natural or synthetic genes and therapeutic genes which are introduced into the subject in the form of DNA or RNA which may function through a DNA intermediate such as by reverse transcriptase. Such genes can be introduced into target cells, directly introduced into the subject, or indirectly introduced by the transfer of transformed cells into the subject. The term “therapeutic gene” means a gene which imparts a beneficial function to the host cell in which such gene is expressed. Therapeutic genes are not naturally found in host cells.

[00185] The term “ecdysone receptor complex” generally refers to a heterodimeric protein complex consisting of two members of the steroid receptor family, ecdysone receptor (“EcR”) and ultraspiracle (“USP”) proteins (see Yao, T.P., et. al. (1993) *Nature* 366, 476-479; Yao, T.-P., et. al., (1992) *Cell* 71, 63-72). The functional ecdysteroid receptor complex may also include additional protein(s) such as immunophilins. Additional members of the steroid receptor family of proteins, known as transcriptional factors (such as DHR38, *betaFTZ-1* or other insect homologs), may also be ligand dependent or independent partners for EcR and/or USP. The ecdysone receptor complex can also be a heterodimer of ecdysone receptor protein and the vertebrate homolog of ultraspiracle protein, retinoic acid-X-receptor (“RXR”) protein. Homodimer complexes of the ecdysone receptor protein or USP may also be functional under some circumstances.

[00186] An ecdysteroid receptor complex can be activated by an active ecdysteroid or non-steroidal ligand bound to one of the proteins of the complex, inclusive of EcR, but not excluding other proteins of the complex.

[00187] The ecdysone receptor complex includes proteins which are members of the steroid receptor superfamily wherein all members are characterized by the presence of an amino-terminal transactivation domain, a DNA binding domain (“DBD”), and a ligand binding domain (“LBD”) separated by a hinge region. Some members of the family may also have another transactivation domain on the carboxy-terminal side of the LBD. The DBD is characterized by the presence of two cysteine zinc fingers between which are two amino acid motifs, the P-box and the D-box, which confer specificity for ecdysone response elements. These domains may be either native, modified, or chimeras of different domains of heterologous receptor proteins.

[00188] The DNA sequences making up the exogenous gene, the response element, and the ecdysone receptor complex may be incorporated into archaeobacteria, procaryotic cells such as *Escherichia coli*, *Bacillus subtilis*, or other enterobacteria, or eucaryotic cells such as plant or animal cells. However, because many of the proteins expressed by the gene are processed incorrectly in

bacteria, eucaryotic cells are preferred. The cells may be in the form of single cells or multicellular organisms. The nucleotide sequences for the exogenous gene, the response element, and the receptor complex can also be incorporated as RNA molecules, preferably in the form of functional viral RNAs such as tobacco mosaic virus. Of the eucaryotic cells, vertebrate cells are preferred because they naturally lack the molecules which confer responses to the ligands of this invention for the ecdysone receptor. As a result, they are insensitive to the ligands of this invention. Thus, the ligands of this invention will have negligible physiological or other effects on transformed cells, or the whole organism. Therefore, cells can grow and express the desired product, substantially unaffected by the presence of the ligand itself.

[00189] The term “subject” means an intact plant or animal or a cell from a plant or animal. It is also anticipated that the ligands will work equally well when the subject is a fungus or yeast. When the subject is an intact animal, preferably the animal is a vertebrate, most preferably a mammal.

[00190] The ligands of the present invention, when used with the ecdysone receptor complex which in turn is bound to the response element linked to an exogenous gene, provide the means for external temporal regulation of expression of the exogenous gene. The order in which the various components bind to each other, that is, ligand to receptor complex and receptor complex to response element, is not critical. Typically, modulation of expression of the exogenous gene is in response to the binding of the ecdysone receptor complex to a specific control, or regulatory, DNA element. The ecdysone receptor protein, like other members of the steroid receptor family, possesses at least three domains, a transactivation domain, a DNA binding domain, and a ligand binding domain. This receptor, like a subset of the steroid receptor family, also possesses less well-defined regions responsible for heterodimerization properties. Binding of the ligand to the ligand binding domain of ecdysone receptor protein, after heterodimerization with USP or RXR protein, enables the DNA binding domains of the heterodimeric proteins to bind to the response element in an activated form, thus resulting in expression or suppression of the exogenous gene. This mechanism does not exclude the potential for ligand binding to either EcR or USP, and the resulting formation of active homodimer complexes (e.g. EcR+EcR or USP+USP). Preferably, one or more of the receptor domains can be varied producing a chimeric gene switch. Typically, one or more of the three domains may be chosen from a source different than the source of the other domains so that the chimeric receptor is optimized in the chosen host cell or organism for transactivating activity, complementary binding of the ligand, and recognition of a specific response element. In addition, the response element itself can be modified or substituted with response elements for other DNA binding protein domains such as the GAL-4 protein from yeast (see Sadowski, et. al. (1988) *Nature*, 335, 563-564) or LexA protein from *E. coli* (see Brent and Ptashne (1985), *Cell*, 43, 729-736) to accommodate chimeric ecdysone receptor complexes. Another advantage of chimeric systems is that they allow choice of a promoter used to drive the exogenous gene according to a desired end result. Such double control can be particularly

important in areas of gene therapy, especially when cytotoxic proteins are produced, because both the timing of expression as well as the cells wherein expression occurs can be controlled. The term “promoter” means a specific nucleotide sequence recognized by RNA polymerase. The sequence is the site at which transcription can be specifically initiated under proper conditions. When exogenous genes, operatively linked to a suitable promoter, are introduced into the cells of the subject, expression of the exogenous genes is controlled by the presence of the ligand of this invention. Promoters may be constitutively or inducibly regulated or may be tissue-specific (that is, expressed only in a particular type of cell) or specific to certain developmental stages of the organism.

[00191] Another aspect of this invention is a method to modulate the expression of one or more exogenous genes in a subject, comprising administering to the subject an effective amount, that is, the amount required to elicit the desired gene expression or suppression, of a ligand comprising a compound of the present invention and wherein the cells of the subject contain:

- a) an ecdysone receptor complex comprising:
 - 1) a DNA binding domain;
 - 2) a binding domain for the ligand; and
 - 3) a transactivation domain; and
- b) a DNA construct comprising:
 - 1) the exogenous gene; and
 - 2) a response element;

wherein the exogenous gene is under the control of the response element; and binding of the DNA binding domain to the response element in the presence of the ligand results in activation or suppression of the gene.

[00192] A related aspect of this invention is a method for regulating endogenous or heterologous gene expression in a transgenic subject comprising contacting a ligand comprising a compound of the present invention with an ecdysone receptor within the cells of the subject wherein the cells contain a DNA binding sequence for the ecdysone receptor and wherein formation of an ecdysone receptor-ligand-DNA binding sequence complex induces expression of the gene.

[00193] A fourth aspect of the present invention is a method for producing a polypeptide comprising the steps of:

- a) selecting a cell which is substantially insensitive to exposure to a ligand comprising a compound of the present invention;
- b) introducing into the cell:
 - 1) a DNA construct comprising:
 - i) an exogenous gene encoding the polypeptide; and
 - ii) a response element;

wherein the gene is under the control of the response element; and

- 2) an ecdysone receptor complex comprising:
 - i) a DNA binding domain;
 - ii) a binding domain for the ligand; and
 - iii) a transactivation domain; and
- c) exposing the cell to the ligand.

[00194] As well as the advantage of temporally controlling polypeptide production by the cell, this aspect of the invention provides a further advantage, in those cases when accumulation of such a polypeptide can damage the cell, in that expression of the polypeptide may be limited to short periods. Such control is particularly important when the exogenous gene is a therapeutic gene. Therapeutic genes may be called upon to produce polypeptides which control needed functions, such as the production of insulin in diabetic patients. They may also be used to produce damaging or even lethal proteins, such as those lethal to cancer cells. Such control may also be important when the protein levels produced may constitute a metabolic drain on growth or reproduction, such as in transgenic plants.

[00195] Numerous genomic and cDNA nucleic acid sequences coding for a variety of polypeptides are well known in the art. Exogenous genetic material useful with the ligands of this invention include genes that encode biologically active proteins of interest, such as, for example, secretory proteins that can be released from a cell; enzymes that can metabolize a substrate from a toxic substance to a non-toxic substance, or from an inactive substance to an active substance; regulatory proteins; cell surface receptors; and the like. Useful genes also include genes that encode blood clotting factors, hormones such as insulin, parathyroid hormone, luteinizing hormone releasing factor, alpha and beta seminal inhibins, and human growth hormone; genes that encode proteins such as enzymes, the absence of which leads to the occurrence of an abnormal state; genes encoding cytokines or lymphokines such as interferons, granulocytic macrophage colony stimulating factor, colony stimulating factor-1, tumor necrosis factor, and erythropoietin; genes encoding inhibitor substances such as alpha₁-antitrypsin, genes encoding substances that function as drugs such as diphtheria and cholera toxins; and the like. Useful genes also include those useful for cancer therapies and to treat genetic disorders. Those skilled in the art have access to nucleic acid sequence information for virtually all known genes and can either obtain the nucleic acid molecule directly from a public depository, the institution that published the sequence, or employ routine methods to prepare the molecule.

[00196] For gene therapy use, the ligands described herein may be taken up in pharmaceutically acceptable carriers, such as, for example, solutions, suspensions, tablets, capsules, ointments, elixirs, and injectable compositions. Pharmaceutical preparations may contain from 0.01 % to 99% by weight of the ligand. Preparations may be either in single or multiple dose forms. The amount of

ligand in any particular pharmaceutical preparation will depend upon the effective dose, that is, the dose required to elicit the desired gene expression or suppression.

[00197] Suitable routes of administering the pharmaceutical preparations include oral, rectal, topical (including dermal, buccal and sublingual), vaginal, parenteral (including subcutaneous, intramuscular, intravenous, intradermal, intrathecal and epidural) and by naso-gastric tube. It will be understood by those skilled in the art that the preferred route of administration will depend upon the condition being treated and may vary with factors such as the condition of the recipient.

[00198] The ligands described herein may also be administered in conjunction with other pharmaceutically active compounds. It will be understood by those skilled in the art that pharmaceutically active compounds to be used in combination with the ligands described herein will be selected in order to avoid adverse effects on the recipient or undesirable interactions between the compounds. Examples of other pharmaceutically active compounds which may be used in combination with the ligands include, for example, AIDS chemotherapeutic agents, amino acid derivatives, analgesics, anesthetics, anorectal products, antacids and antiflatulents, antibiotics, anticoagulants, antidotes, antifibrinolytic agents, antihistamines, anti-inflammatory agents, antineoplastics, antiparasitics, antiprotozoals, antipyretics, antiseptics, antispasmodics and anticholinergics, antivirals, appetite suppressants, arthritis medications, biological response modifiers, bone metabolism regulators, bowel evacuants, cardiovascular agents, central nervous system stimulants, cerebral metabolic enhancers, cerumenolytics, cholinesterase inhibitors, cold and cough preparations, colony stimulating factors, contraceptives, cytoprotective agents, dental preparations, deodorants, dermatologicals, detoxifying agents, diabetes agents, diagnostics, diarrhea medications, dopamine receptor agonists, electrolytes, enzymes and digestants, ergot preparations, fertility agents, fiber supplements, antifungal agents, galactorrhea inhibitors, gastric acid secretion inhibitors, gastrointestinal prokinetic agents, gonadotropin inhibitors, hair growth stimulants, hematinics, hemorrhheologic agents, hemostatics, histamine H₂ receptor antagonists, hormones, hyperglycemic agents, hypolipidemics, immunosuppressants, laxatives, leprostatics, leukapheresis adjuncts, lung surfactants, migraine preparations, mucolytics, muscle relaxant antagonists, muscle relaxants, narcotic antagonists, nasal sprays, nausea medications nucleoside analogues, nutritional supplements, osteoporosis preparations, oxytocics, parasympatholytics, parasympathomimetics, Parkinsonism drugs, Penicillin adjuvants, phospholipids, platelet inhibitors, porphyria agents, prostaglandin analogues, prostaglandins, proton pump inhibitors, pruritus medications psychotropics, quinolones, respiratory stimulants, saliva stimulants, salt substitutes, sclerosing agents, skin wound preparations, smoking cessation aids, sulfonamides, sympatholytics, thrombolytics, Tourette's syndrome agents, tremor preparations, tuberculosis preparations, uricosuric agents, urinary tract agents, uterine contractants, uterine relaxants, vaginal preparations, vertigo agents, vitamin D analogs, vitamins, and

medical imaging contrast media. In some cases the ligands may be useful as an adjunct to drug therapy, for example, to "turn off" a gene that produces an enzyme that metabolizes a particular drug. [00199] For agricultural applications, in addition to the applications described above, the ligands of this invention may also be used to control the expression of pesticidal proteins such as *Bacillus thuringiensis* (Bt) toxin. Such expression may be tissue or plant specific. In addition, particularly when control of plant pests is also needed, one or more pesticides may be combined with the ligands described herein, thereby providing additional advantages and effectiveness, including fewer total applications, than if the pesticides are applied separately. When mixtures with pesticides are employed, the relative proportions of each component in the composition will depend upon the relative efficacy and the desired application rate of each pesticide with respect to the crops, pests, and/or weeds to be treated. Those skilled in the art will recognize that mixtures of pesticides may provide advantages such as a broader spectrum of activity than one pesticide used alone. Examples of pesticides which can be combined in compositions with the ligands described herein include fungicides, herbicides, insecticides, miticides, and microbicides.

[00200] The ligands described herein can be applied to plant foliage as aqueous sprays by methods commonly employed, such as conventional high-liter hydraulic sprays, low-liter sprays, air-blast, and aerial sprays. The dilution and rate of application will depend upon the type of equipment employed, the method and frequency of application desired, and the ligand application rate. It may be desirable to include additional adjuvants in the spray tank. Such adjuvants include surfactants, dispersants, spreaders, stickers, antifoam agents, emulsifiers, and other similar materials described in *McCutcheon's Emulsifiers and Detergents*, *McCutcheon's Emulsifiers and Detergents/Functional Materials*, and *McCutcheon's Functional Materials*, all published annually by McCutcheon Division of MC Publishing Company (New Jersey). The ligands can also be mixed with fertilizers or fertilizing materials before their application. The ligands and solid fertilizing material can also be admixed in mixing or blending equipment, or they can be incorporated with fertilizers in granular formulations. Any relative proportion of fertilizer can be used which is suitable for the crops and weeds to be treated. The ligands described herein will commonly comprise from 5% to 50% of the fertilizing composition. These compositions provide fertilizing materials which promote the rapid growth of desired plants, and at the same time control gene expression.

HOST CELLS AND NON-HUMAN ORGANISMS OF THE INVENTION

[00201] As described above, ligands for modulating gene expression system of the present invention may be used to modulate gene expression in a host cell. Expression in transgenic host cells may be useful for the expression of various genes of interest. The present invention provides ligands for modulation of gene expression in prokaryotic and eukaryotic host cells. Expression in transgenic host cells is useful for the expression of various polypeptides of interest including but not limited to

antigens produced in plants as vaccines, enzymes like alpha-amylase, phytase, glucanase, and xylanase, genes for resistance against insects, nematodes, fungi, bacteria, viruses, and abiotic stresses, antigens, nutraceuticals, pharmaceuticals, vitamins, genes for modifying amino acid content, herbicide resistance, cold, drought, and heat tolerance, industrial products, oils, protein, carbohydrates, antioxidants, male sterile plants, flowers, fuels, other output traits, therapeutic polypeptides, pathway intermediates; for the modulation of pathways already existing in the host for the synthesis of new products heretofore not possible using the host; cell based assays; functional genomics assays, biotherapeutic protein production, proteomics assays, and the like. Additionally the gene products may be useful for conferring higher growth yields of the host or for enabling an alternative growth mode to be utilized.

[00202] Thus, the present invention provides ligands for modulating gene expression in an isolated host cell according to the invention. The host cell may be a bacterial cell, a fungal cell, a nematode cell, an insect cell, a fish cell, a plant cell, an avian cell, an animal cell, or a mammalian cell. In still another embodiment, the invention relates to ligands for modulating gene expression in an host cell, wherein the method comprises culturing the host cell as described above in culture medium under conditions permitting expression of a polynucleotide encoding the nuclear receptor ligand binding domain comprising a substitution mutation, and isolating the nuclear receptor ligand binding domain comprising a substitution mutation from the culture.

[00203] In a specific embodiment, the isolated host cell is a prokaryotic host cell or a eukaryotic host cell. In another specific embodiment, the isolated host cell is an invertebrate host cell or a vertebrate host cell. Preferably, the host cell is selected from the group consisting of a bacterial cell, a fungal cell, a yeast cell, a nematode cell, an insect cell, a fish cell, a plant cell, an avian cell, an animal cell, and a mammalian cell. More preferably, the host cell is a yeast cell, a nematode cell, an insect cell, a plant cell, a zebrafish cell, a chicken cell, a hamster cell, a mouse cell, a rat cell, a rabbit cell, a cat cell, a dog cell, a bovine cell, a goat cell, a cow cell, a pig cell, a horse cell, a sheep cell, a simian cell, a monkey cell, a chimpanzee cell, or a human cell. Examples of preferred host cells include, but are not limited to, fungal or yeast species such as *Aspergillus*, *Trichoderma*, *Saccharomyces*, *Pichia*, *Candida*, *Hansenula*, or bacterial species such as those in the genera *Synechocystis*, *Synechococcus*, *Salmonella*, *Bacillus*, *Acinetobacter*, *Rhodococcus*, *Streptomyces*, *Escherichia*, *Pseudomonas*, *Methylobacter*, *Alcaligenes*, *Synechocystis*, *Anabaena*, *Thiobacillus*, *Methanobacterium* and *Klebsiella*; plant species selected from the group consisting of an apple, *Arabidopsis*, bajra, banana, barley, beans, beet, blackgram, chickpea, chili, cucumber, eggplant, favabean, maize, melon, millet, mungbean, oat, okra, *Panicum*, papaya, peanut, pea, pepper, pigeonpea, pineapple, *Phaseolus*, potato, pumpkin, rice, sorghum, soybean, squash, sugarcane, sugarbeet, sunflower, sweet potato, tea, tomato, tobacco, watermelon, and wheat; animal; and mammalian host cells.

[00204] In a specific embodiment, the host cell is a yeast cell selected from the group consisting of a *Saccharomyces*, a *Pichia*, and a *Candida* host cell.

[00205] In another specific embodiment, the host cell is a *Caenorhabdus elegans* nematode cell.

[00206] In another specific embodiment, the host cell is an insect cell.

[00207] In another specific embodiment, the host cell is a plant cell selected from the group consisting of an apple, *Arabidopsis*, bajra, banana, barley, beans, beet, blackgram, chickpea, chili, cucumber, eggplant, favabean, maize, melon, millet, mungbean, oat, okra, *Panicum*, papaya, peanut, pea, pepper, pigeonpea, pineapple, *Phaseolus*, potato, pumpkin, rice, sorghum, soybean, squash, sugarcane, sugarbeet, sunflower, sweet potato, tea, tomato, tobacco, watermelon, and wheat cell.

[00208] In another specific embodiment, the host cell is a zebrafish cell.

[00209] In another specific embodiment, the host cell is a chicken cell.

[00210] In another specific embodiment, the host cell is a mammalian cell selected from the group consisting of a hamster cell, a mouse cell, a rat cell, a rabbit cell, a cat cell, a dog cell, a bovine cell, a goat cell, a cow cell, a pig cell, a horse cell, a sheep cell, a monkey cell, a chimpanzee cell, and a human cell.

[00211] Host cell transformation is well known in the art and may be achieved by a variety of methods including but not limited to electroporation, viral infection, plasmid/vector transfection, non-viral vector mediated transfection, *Agrobacterium*-mediated transformation, particle bombardment, and the like. Expression of desired gene products involves culturing the transformed host cells under suitable conditions and inducing expression of the transformed gene. Culture conditions and gene expression protocols in prokaryotic and eukaryotic cells are well known in the art (see General Methods section of Examples). Cells may be harvested and the gene products isolated according to protocols specific for the gene product.

[00212] In addition, a host cell may be chosen which modulates the expression of the inserted polynucleotide, or modifies and processes the polypeptide product in the specific fashion desired. Different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification [*e.g.*, glycosylation, cleavage (*e.g.*, of signal sequence)] of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce a non-glycosylated core protein product. However, a polypeptide expressed in bacteria may not be properly folded. Expression in yeast can produce a glycosylated product. Expression in eukaryotic cells can increase the likelihood of "native" glycosylation and folding of a heterologous protein. Moreover, expression in mammalian cells can provide a tool for reconstituting, or constituting, the polypeptide's activity. Furthermore, different vector/host expression systems may affect processing reactions, such as proteolytic cleavages, to a different extent. The present invention also relates to a non-human organism comprising an isolated host cell according to the invention. In a

specific embodiment, the non-human organism is a prokaryotic organism or a eukaryotic organism. In another specific embodiment, the non-human organism is an invertebrate organism or a vertebrate organism.

[00213] Preferably, the non-human organism is selected from the group consisting of a bacterium, a fungus, a yeast, a nematode, an insect, a fish, a plant, a bird, an animal, and a mammal. More preferably, the non-human organism is a yeast, a nematode, an insect, a plant, a zebrafish, a chicken, a hamster, a mouse, a rat, a rabbit, a cat, a dog, a bovine, a goat, a cow, a pig, a horse, a sheep, a simian, a monkey, or a chimpanzee.

[00214] In a specific embodiment, the non-human organism is a yeast selected from the group consisting of *Saccharomyces*, *Pichia*, and *Candida*.

[00215] In another specific embodiment, the non-human organism is a *Caenorhabditis elegans* nematode.

[00216] In another specific embodiment, the non-human organism is a plant selected from the group consisting of an apple, *Arabidopsis*, bajra, banana, barley, beans, beet, blackgram, chickpea, chili, cucumber, eggplant, favabean, maize, melon, millet, mungbean, oat, okra, *Panicum*, papaya, peanut, pea, pepper, pigeonpea, pineapple, *Phaseolus*, potato, pumpkin, rice, sorghum, soybean, squash, sugarcane, sugarbeet, sunflower, sweet potato, tea, tomato, tobacco, watermelon, and wheat.

[00217] In another specific embodiment, the non-human organism is a *Mus musculus* mouse.

GENE EXPRESSION MODULATION SYSTEM OF THE INVENTION

[00218] The present invention relates to a group of ligands that are useful in an ecdysone receptor-based inducible gene expression system. As presented herein, a novel group of ligands provides an improved inducible gene expression system in both prokaryotic and eukaryotic host cells. Thus, the present invention relates to ligands that are useful to modulate expression of genes. In particular, the present invention relates to ligands having the ability to transactivate a gene expression modulation system comprising at least one gene expression cassette that is capable of being expressed in a host cell comprising a polynucleotide that encodes a polypeptide comprising a Group H nuclear receptor ligand binding domain. Preferably, the Group H nuclear receptor ligand binding is from an ecdysone receptor, a ubiquitous receptor, an orphan receptor 1, a NER-1, a steroid hormone nuclear receptor 1, a retinoid X receptor interacting protein-15, a liver X receptor β , a steroid hormone receptor like protein, a liver X receptor, a liver X receptor α , a farnesoid X receptor, a receptor interacting protein 14, and a farnesol receptor. More preferably, the Group H nuclear receptor ligand binding domain is from an ecdysone receptor.

[00219] In a specific embodiment, the gene expression modulation system comprises a gene expression cassette comprising a polynucleotide that encodes a polypeptide comprising a transactivation domain, a DNA-binding domain that recognizes a response element associated with a

gene whose expression is to be modulated; and a Group H nuclear receptor ligand binding domain comprising a substitution mutation. The gene expression modulation system may further comprise a second gene expression cassette comprising: i) a response element recognized by the DNA-binding domain of the encoded polypeptide of the first gene expression cassette; ii) a promoter that is activated by the transactivation domain of the encoded polypeptide of the first gene expression cassette; and iii) a gene whose expression is to be modulated.

[00220] In another specific embodiment, the gene expression modulation system comprises a gene expression cassette comprising a) a polynucleotide that encodes a polypeptide comprising a transactivation domain, a DNA-binding domain that recognizes a response element associated with a gene whose expression is to be modulated; and a Group H nuclear receptor ligand binding domain comprising a substitution mutation, and b) a second nuclear receptor ligand binding domain selected from the group consisting of a vertebrate retinoid X receptor ligand binding domain, an invertebrate retinoid X receptor ligand binding domain, an ultraspiracle protein ligand binding domain, and a chimeric ligand binding domain comprising two polypeptide fragments, wherein the first polypeptide fragment is from a vertebrate retinoid X receptor ligand binding domain, an invertebrate retinoid X receptor ligand binding domain, or an ultraspiracle protein ligand binding domain, and the second polypeptide fragment is from a different vertebrate retinoid X receptor ligand binding domain, invertebrate retinoid X receptor ligand binding domain, or ultraspiracle protein ligand binding domain. The gene expression modulation system may further comprise a second gene expression cassette comprising: i) a response element recognized by the DNA-binding domain of the encoded polypeptide of the first gene expression cassette; ii) a promoter that is activated by the transactivation domain of the encoded polypeptide of the first gene expression cassette; and iii) a gene whose expression is to be modulated.

[00221] In another specific embodiment, the gene expression modulation system comprises a first gene expression cassette comprising a polynucleotide that encodes a first polypeptide comprising a DNA-binding domain that recognizes a response element associated with a gene whose expression is to be modulated and a nuclear receptor ligand binding domain, and a second gene expression cassette comprising a polynucleotide that encodes a second polypeptide comprising a transactivation domain and a nuclear receptor ligand binding domain, wherein one of the nuclear receptor ligand binding domains is a Group H nuclear receptor ligand binding domain comprising a substitution mutation. In a preferred embodiment, the first polypeptide is substantially free of a transactivation domain and the second polypeptide is substantially free of a DNA binding domain. For purposes of the invention, "substantially free" means that the protein in question does not contain a sufficient sequence of the domain in question to provide activation or binding activity. The gene expression modulation system may further comprise a third gene expression cassette comprising: i) a response element recognized by the DNA-binding domain of the first polypeptide of the first gene expression cassette; ii) a

promoter that is activated by the transactivation domain of the second polypeptide of the second gene expression cassette; and iii) a gene whose expression is to be modulated.

[00222] Wherein when only one nuclear receptor ligand binding domain is a Group H ligand binding domain comprising a substitution mutation, the other nuclear receptor ligand binding domain may be from any other nuclear receptor that forms a dimer with the Group H ligand binding domain comprising the substitution mutation. For example, when the Group H nuclear receptor ligand binding domain comprising a substitution mutation is an ecdysone receptor ligand binding domain comprising a substitution mutation, the other nuclear receptor ligand binding domain (“partner”) may be from an ecdysone receptor, a vertebrate retinoid X receptor (RXR), an invertebrate RXR, an ultraspiracle protein (USP), or a chimeric nuclear receptor comprising at least two different nuclear receptor ligand binding domain polypeptide fragments selected from the group consisting of a vertebrate RXR, an invertebrate RXR, and a USP (see co-pending applications PCT/US01/09050, PCT/US02/05235, and PCT/US02/05706, incorporated herein by reference in their entirety). The “partner” nuclear receptor ligand binding domain may further comprise a truncation mutation, a deletion mutation, a substitution mutation, or another modification.

[00223] Preferably, the vertebrate RXR ligand binding domain is from a human *Homo sapiens*, mouse *Mus musculus*, rat *Rattus norvegicus*, chicken *Gallus gallus*, pig *Sus scrofa domestica*, frog *Xenopus laevis*, zebrafish *Danio rerio*, tunicate *Polyandrocarpa misakiensis*, or jellyfish *Tripedalia cysophora* RXR.

[00224] Preferably, the invertebrate RXR ligand binding domain is from a locust *Locusta migratoria* ultraspiracle polypeptide (“LmUSP”), an ixodid tick *Amblyomma americanum* RXR homolog 1 (“AmaRXR1”), a ixodid tick *Amblyomma americanum* RXR homolog 2 (“AmaRXR2”), a fiddler crab *Celuca pugilator* RXR homolog (“CpRXR”), a beetle *Tenebrio molitor* RXR homolog (“TmRXR”), a honeybee *Apis mellifera* RXR homolog (“AmRXR”), an aphid *Myzus persicae* RXR homolog (“MpRXR”), or a non-Dipteran/non-Lepidopteran RXR homolog.

[00225] Preferably, the chimeric RXR ligand binding domain comprises at least two polypeptide fragments selected from the group consisting of a vertebrate species RXR polypeptide fragment, an invertebrate species RXR polypeptide fragment, and a non-Dipteran/non-Lepidopteran invertebrate species RXR homolog polypeptide fragment. A chimeric RXR ligand binding domain for use in the present invention may comprise at least two different species RXR polypeptide fragments, or when the species is the same, the two or more polypeptide fragments may be from two or more different isoforms of the species RXR polypeptide fragment.

[00226] In a preferred embodiment, the chimeric RXR ligand binding domain comprises at least one vertebrate species RXR polypeptide fragment and one invertebrate species RXR polypeptide fragment.

[00227] In a more preferred embodiment, the chimeric RXR ligand binding domain comprises at least one vertebrate species RXR polypeptide fragment and one non-Dipteran/non-Lepidopteran invertebrate species RXR homolog polypeptide fragment.

[00228] In a specific embodiment, the gene whose expression is to be modulated is a homologous gene with respect to the host cell. In another specific embodiment, the gene whose expression is to be modulated is a heterologous gene with respect to the host cell.

[00229] The ligands for use in the present invention as described below, when combined with the ligand binding domain of the nuclear receptor(s), which in turn are bound to the response element linked to a gene, provide the means for external temporal regulation of expression of the gene. The binding mechanism or the order in which the various components of this invention bind to each other, that is, for example, ligand to ligand binding domain, DNA-binding domain to response element, transactivation domain to promoter, etc., is not critical.

[00230] In a specific example, binding of the ligand to the ligand binding domain of a Group H nuclear receptor and its nuclear receptor ligand binding domain partner enables expression or suppression of the gene. This mechanism does not exclude the potential for ligand binding to the Group H nuclear receptor (GHNR) or its partner, and the resulting formation of active homodimer complexes (e.g. GHNR + GHNR or partner + partner). Preferably, one or more of the receptor domains is varied producing a hybrid gene switch. Typically, one or more of the three domains, DBD, LBD, and transactivation domain, may be chosen from a source different than the source of the other domains so that the hybrid genes and the resulting hybrid proteins are optimized in the chosen host cell or organism for transactivating activity, complementary binding of the ligand, and recognition of a specific response element. In addition, the response element itself can be modified or substituted with response elements for other DNA binding protein domains such as the GAL-4 protein from yeast (see Sadowski, et al. (1988) *Nature*, 335: 563-564) or LexA protein from *Escherichia coli* (see Brent and Ptashne (1985), *Cell*, 43: 729-736), or synthetic response elements specific for targeted interactions with proteins designed, modified, and selected for such specific interactions (see, for example, Kim, et al. (1997), *Proc. Natl. Acad. Sci., USA*, 94: 3616-3620) to accommodate hybrid receptors. Another advantage of two-hybrid systems is that they allow choice of a promoter used to drive the gene expression according to a desired end result. Such double control can be particularly important in areas of gene therapy, especially when cytotoxic proteins are produced, because both the timing of expression as well as the cells wherein expression occurs can be controlled. When genes, operably linked to a suitable promoter, are introduced into the cells of the subject, expression of the exogenous genes is controlled by the presence of the system of this invention. Promoters may be constitutively or inducibly regulated or may be tissue-specific (that is, expressed only in a particular type of cells) or specific to certain developmental stages of the organism.

[00231] The ecdysone receptor is a member of the nuclear receptor superfamily and classified into subfamily 1, group H (referred to herein as “Group H nuclear receptors”). The members of each group share 40-60% amino acid identity in the E (ligand binding) domain (Laudet et al., A Unified Nomenclature System for the Nuclear Receptor Subfamily, 1999; Cell 97: 161-163). In addition to the ecdysone receptor, other members of this nuclear receptor subfamily 1, group H include: ubiquitous receptor (UR), orphan receptor 1 (OR-1), steroid hormone nuclear receptor 1 (NER-1), retinoid X receptor interacting protein -15 (RIP-15), liver X receptor β (LXR β), steroid hormone receptor like protein (RLD-1), liver X receptor (LXR), liver X receptor α (LXR α), farnesoid X receptor (FXR), receptor interacting protein 14 (RIP-14), and farnesol receptor (HRR-1

[00232] In particular, described herein are novel ligands useful in a gene expression modulation system comprising a Group H nuclear receptor ligand binding domain comprising a substitution mutation. This gene expression system may be a “single switch”-based gene expression system in which the transactivation domain, DNA-binding domain and ligand binding domain are on one encoded polypeptide. Alternatively, the gene expression modulation system may be a “dual switch”- or “two-hybrid”-based gene expression modulation system in which the transactivation domain and DNA-binding domain are located on two different encoded polypeptides.

[00233] An ecdysone receptor-based gene expression modulation system of the present invention may be either heterodimeric or homodimeric. A functional EcR complex generally refers to a heterodimeric protein complex consisting of two members of the steroid receptor family, an ecdysone receptor protein obtained from various insects, and an ultraspiracle (USP) protein or the vertebrate homolog of USP, retinoid X receptor protein (see Yao, et al. (1993) Nature 366, 476-479; Yao, et al., (1992) Cell 71, 63-72). However, the complex may also be a homodimer as detailed below. The functional ecdysteroid receptor complex may also include additional protein(s) such as immunophilins. Additional members of the steroid receptor family of proteins, known as transcriptional factors (such as DHR38 or *betaFTZ-1*), may also be ligand dependent or independent partners for EcR, USP, and/or RXR. Additionally, other cofactors may be required such as proteins generally known as coactivators (also termed adapters or mediators). These proteins do not bind sequence-specifically to DNA and are not involved in basal transcription. They may exert their effect on transcription activation through various mechanisms, including stimulation of DNA-binding of activators, by affecting chromatin structure, or by mediating activator-initiation complex interactions. Examples of such coactivators include RIP140, TIF1, RAP46/Bag-1, ARA70, SRC-1/NCoA-1, TIF2/GRIP/NCoA-2, ACTR/AIB1/RAC3/pCIP as well as the promiscuous coactivator C response element B binding protein, CBP/p300 (for review see Glass et al., Curr. Opin. Cell Biol. 9:222-232, 1997). Also, protein cofactors generally known as corepressors (also known as repressors, silencers, or silencing mediators) may be required to effectively inhibit transcriptional activation in the absence of ligand. These corepressors may interact with the unliganded ecdysone receptor to silence the

activity at the response element. Current evidence suggests that the binding of ligand changes the conformation of the receptor, which results in release of the corepressor and recruitment of the above described coactivators, thereby abolishing their silencing activity. Examples of corepressors include N-CoR and SMRT (for review, see Horwitz et al. *Mol Endocrinol.* 10: 1167-1177, 1996). These cofactors may either be endogenous within the cell or organism, or may be added exogenously as transgenes to be expressed in either a regulated or unregulated fashion. Homodimer complexes of the ecdysone receptor protein, USP, or RXR may also be functional under some circumstances.

[00234] The ecdysone receptor complex typically includes proteins that are members of the nuclear receptor superfamily wherein all members are generally characterized by the presence of an amino-terminal transactivation domain, a DNA binding domain (“DBD”), and a ligand binding domain (“LBD”) separated from the DBD by a hinge region. As used herein, the term “DNA binding domain” comprises a minimal polypeptide sequence of a DNA binding protein, up to the entire length of a DNA binding protein, so long as the DNA binding domain functions to associate with a particular response element. Members of the nuclear receptor superfamily are also characterized by the presence of four or five domains: A/B, C, D, E, and in some members F (see US patent 4,981,784 and Evans, *Science* 240:889-895 (1988)). The “A/B” domain corresponds to the transactivation domain, “C” corresponds to the DNA binding domain, “D” corresponds to the hinge region, and “E” corresponds to the ligand binding domain. Some members of the family may also have another transactivation domain on the carboxy-terminal side of the LBD corresponding to “F”.

[00235] The DBD is characterized by the presence of two cysteine zinc fingers between which are two amino acid motifs, the P-box and the D-box, which confer specificity for ecdysone response elements. These domains may be either native, modified, or chimeras of different domains of heterologous receptor proteins. The EcR receptor, like a subset of the steroid receptor family, also possesses less well-defined regions responsible for heterodimerization properties. Because the domains of nuclear receptors are modular in nature, the LBD, DBD, and transactivation domains may be interchanged.

[00236] Gene switch systems are known that incorporate components from the ecdysone receptor complex. However, in these known systems, whenever EcR is used it is associated with native or modified DNA binding domains and transactivation domains on the same molecule. USP or RXR are typically used as silent partners. It has previously been shown that when DNA binding domains and transactivation domains are on the same molecule the background activity in the absence of ligand is high and that such activity is dramatically reduced when DNA binding domains and transactivation domains are on different molecules, that is, on each of two partners of a heterodimeric or homodimeric complex (see PCT/US01/09050).

METHOD OF MODULATING GENE EXPRESSION OF THE INVENTION

[00237] The present invention also relates to methods of modulating gene expression in a host cell using a gene expression modulation system according to the invention. Specifically, the present invention provides a method of modulating the expression of a gene in a host cell comprising the steps of: a) introducing into the host cell a gene expression modulation system according to the invention; and b) introducing into the host cell a ligand; wherein the gene to be modulated is a component of a gene expression cassette comprising: i) a response element comprising a domain recognized by the DNA binding domain of the gene expression system; ii) a promoter that is activated by the transactivation domain of the gene expression system; and iii) a gene whose expression is to be modulated, whereby upon introduction of the ligand into the host cell, expression of the gene is modulated.

[00238] The invention also provides a method of modulating the expression of a gene in a host cell comprising the steps of: a) introducing into the host cell a gene expression modulation system according to the invention; b) introducing into the host cell a gene expression cassette according to the invention, wherein the gene expression cassette comprises i) a response element comprising a domain recognized by the DNA binding domain from the gene expression system; ii) a promoter that is activated by the transactivation domain of the gene expression system; and iii) a gene whose expression is to be modulated; and c) introducing into the host cell a ligand; whereby upon introduction of the ligand into the host cell, expression of the gene is modulated.

[00239] The present invention also provides a method of modulating the expression of a gene in a host cell comprising a gene expression cassette comprising a response element comprising a domain to which the DNA binding domain from the first hybrid polypeptide of the gene expression modulation system binds; a promoter that is activated by the transactivation domain of the second hybrid polypeptide of the gene expression modulation system; and a gene whose expression is to be modulated; wherein the method comprises the steps of: a) introducing into the host cell a gene expression modulation system according to the invention; and b) introducing into the host cell a ligand; whereby upon introduction of the ligand into the host, expression of the gene is modulated.

[00240] Genes of interest for expression in a host cell using methods disclosed herein may be endogenous genes or heterologous genes. Nucleic acid or amino acid sequence information for a desired gene or protein can be located in one of many public access databases, for example, GENBANK, EMBL, Swiss-Prot, and PIR, or in many biology related journal publications. Thus, those skilled in the art have access to nucleic acid sequence information for virtually all known genes. Such information can then be used to construct the desired constructs for the insertion of the gene of interest within the gene expression cassettes used in the methods described herein.

[00241] Examples of genes of interest for expression in a host cell using methods set forth herein include, but are not limited to: antigens produced in plants as vaccines, enzymes like alpha-amylase,

phytase, glucanase, and xylanase, genes for resistance against insects, nematodes, fungi, bacteria, viruses, and abiotic stresses, nutraceuticals, pharmaceuticals, vitamins, genes for modifying amino acid content, herbicide resistance, cold, drought, and heat tolerance, industrial products, oils, protein, carbohydrates, antioxidants, male sterile plants, flowers, fuels, other output traits, genes encoding therapeutically desirable polypeptides or products that may be used to treat a condition, a disease, a disorder, a dysfunction, a genetic defect, such as monoclonal antibodies, enzymes, proteases, cytokines, interferons, insulin, erythropoietin, clotting factors, other blood factors or components, viral vectors for gene therapy, virus for vaccines, targets for drug discovery, functional genomics, and proteomics analyses and applications, and the like.

MEASURING GENE EXPRESSION/TRANSCRIPTION

[00242] One useful measurement of the methods of the invention is that of the transcriptional state of the cell including the identities and abundances of RNA, preferably mRNA species. Such measurements are conveniently conducted by measuring cDNA abundances by any of several existing gene expression technologies.

[00243] Nucleic acid array technology is a useful technique for determining differential mRNA expression. Such technology includes, for example, oligonucleotide chips and DNA microarrays. These techniques rely on DNA fragments or oligonucleotides which correspond to different genes or cDNAs which are immobilized on a solid support and hybridized to probes prepared from total mRNA pools extracted from cells, tissues, or whole organisms and converted to cDNA.

Oligonucleotide chips are arrays of oligonucleotides synthesized on a substrate using photolithographic techniques. Chips have been produced which can analyze for up to 1700 genes. DNA microarrays are arrays of DNA samples, typically PCR products, that are robotically printed onto a microscope slide. Each gene is analyzed by a full or partial-length target DNA sequence. Microarrays with up to 10,000 genes are now routinely prepared commercially. The primary difference between these two techniques is that oligonucleotide chips typically utilize 25-mer oligonucleotides which allow fractionation of short DNA molecules whereas the larger DNA targets of microarrays, approximately 1000 base pairs, may provide more sensitivity in fractionating complex DNA mixtures.

[00244] Another useful measurement of the methods of the invention is that of determining the translation state of the cell by measuring the abundances of the constituent protein species present in the cell using processes well known in the art.

[00245] Where identification of genes associated with various physiological functions is desired, an assay may be employed in which changes in such functions as cell growth, apoptosis, senescence, differentiation, adhesion, binding to a specific molecules, binding to another cell, cellular

organization, organogenesis, intracellular transport, transport facilitation, energy conversion, metabolism, myogenesis, neurogenesis, and/or hematopoiesis is measured.

[00246] In addition, selectable marker or reporter gene expression may be used to measure gene expression modulation using the present invention.

[00247] Other methods to detect the products of gene expression are well known in the art and include Southern blots (DNA detection), dot or slot blots (DNA, RNA), northern blots (RNA), RT-PCR (RNA), western blots (polypeptide detection), and ELISA (polypeptide) analyses. Although less preferred, labeled proteins can be used to detect a particular nucleic acid sequence to which it hybridizes.

[00248] In some cases it is necessary to amplify the amount of a nucleic acid sequence. This may be carried out using one or more of a number of suitable methods including, for example, polymerase chain reaction ("PCR"), ligase chain reaction ("LCR"), strand displacement amplification ("SDA"), transcription-based amplification, and the like. PCR is carried out in accordance with known techniques in which, for example, a nucleic acid sample is treated in the presence of a heat stable DNA polymerase, under hybridizing conditions, with one pair of oligonucleotide primers, with one primer hybridizing to one strand (template) of the specific sequence to be detected. The primers are sufficiently complementary to each template strand of the specific sequence to hybridize therewith. An extension product of each primer is synthesized and is complementary to the nucleic acid template strand to which it hybridized. The extension product synthesized from each primer can also serve as a template for further synthesis of extension products using the same primers. Following a sufficient number of rounds of synthesis of extension products, the sample may be analyzed as described above to assess whether the sequence or sequences to be detected are present.

[00249] The present invention may be better understood by reference to the following non-limiting Examples, which are provided as exemplary of the invention.

EXAMPLES

GENERAL METHODS

[00250] Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. (1989) (Maniatis) and by T. J. Silhavy, M. L. Bannan, and L. W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984) and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, Greene Publishing Assoc. and Wiley-Interscience (1987).

[00251] Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out

in *Manual of Methods for General Bacteriology* (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC. (1994)) or by Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of host cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

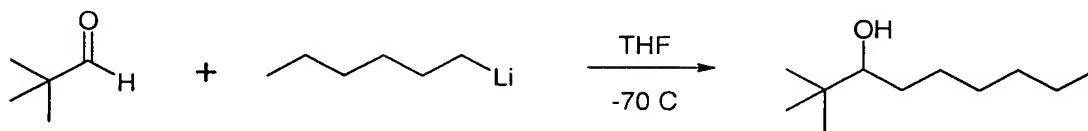
[00252] Manipulations of genetic sequences may be accomplished using the suite of programs available from the Genetics Computer Group Inc. (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI). Where the GCG program "Pileup" is used the gap creation default value of 12, and the gap extension default value of 4 may be used. Where the CGC "Gap" or "Bestfit" program is used the default gap creation penalty of 50 and the default gap extension penalty of 3 may be used. In any case where GCG program parameters are not prompted for, in these or any other GCG program, default values may be used.

[00253] The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "μL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "μM" means micromolar, "mM" means millimolar, "M" means molar, "mmol" means millimoles, "μg" means microgram(s), "mg" means milligram(s), "A" means adenine or adenosine, "T" means thymine or thymidine, "G" means guanine or guanosine, "C" means cytidine or cytosine, "x g" means times gravity, "nt" means nucleotide(s), "aa" means amino acid(s), "bp" means base pair(s), "kb" means kilobase(s), "k" means kilo, "μ" means micro, "°C" means degrees Celsius, "C" in the context of a chemical equation means Celsius, "THF" means tetrahydrofuran, "DME" means dimethoxyethane, "DMF" means dimethylformamide, "NMR" means nuclear magnetic resonance, "psi" refers to pounds per square inch, and "TLC" means thin layer chromatography.

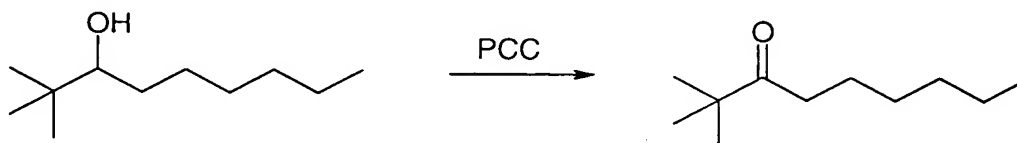
EXAMPLE 1: PREPARATION OF COMPOUNDS

[00254] The compounds of the present invention may be made according to the following synthesis routes.

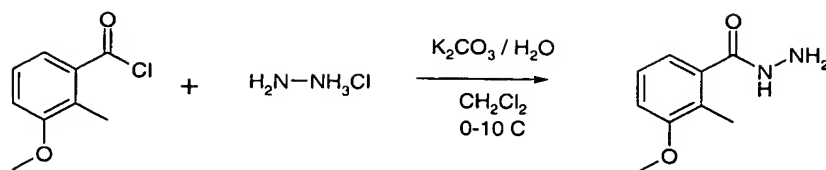
1.1 Preparation of RG-115853



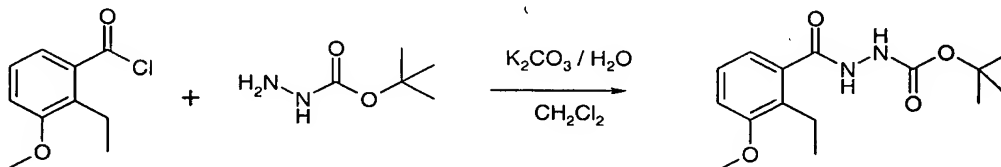
[00255] 20.0 g (0.232 mol) of pivaldehyde were dissolved in 600 mL THF in a 3-neck round bottom 2L flask equipped with a magnetic stir bar and thermometer. The flask was flushed with N₂. The solution was cooled to -65 °C in a dry ice/acetone bath. 100 mL of a 2.3 M solution of hexyllithium in hexane (0.23 moles) was added by means of a 20 mL glass syringe inserted between a rubber stopper and the glass neck, in small 5 mL portions, keeping the temperature at or below -60 °C. After stirring at -60 °C for 1 hour, the reaction was allowed to warm to ca. -5 °C over one hour. The reaction was cooled again to -60 °C, and slowly quenched with aqueous NH₄Cl solution, and allowing the temperature to rise above -50 °C. The reaction mixture was allowed to warm up as 100 mL of water were added. The THF was removed on a rotary evaporator, maintaining the water bath temperature at 25-30 °C, to prevent loss of volatile product. The product was extracted with ethyl ether; the organic layer was dried and solvent was carefully removed on a rotary evaporator, monitoring weight loss. The product 2,2-dimethyl-nonan-3-ol was used in the subsequent oxidation step as a highly concentrated solution in ether.



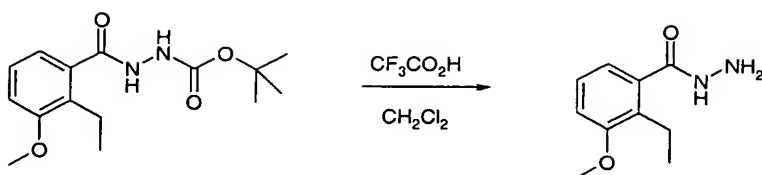
[00256] 2,2-dimethyl-nonan-3-ol, available as a concentrated solution in ether (ca. 0.23 moles, cf. previous procedure), was dissolved in ca. 350 mL CH₂Cl₂ in a 500 mL round bottom flask. With vigorous stirring and external cooling, pyridinium chlorochromate (PCC, 76.6 g, 0.355 mol) was added slowly. The reaction mixture turned black and began to warm. The reaction was stirred overnight at room temperature, and the supernatant was decanted from a black sludge which had formed. The sludge was extracted with ca. 40 mL of hexane, which was combined with the CH₂Cl₂ solution. This mixture was applied directly to a 100 g silica gel column, and eluted first with CH₂Cl₂/hexane and then with 10% ethyl acetate in hexane to yield 35.8 g crude, green-colored product. This material was rechromatographed; elution with hexane and careful evaporation at 25 °C on a rotary evaporator yielded 26.1 g 2,2-dimethyl-nonan-3-one (67% yield). ¹H NMR (500 MHz, CDCl₃), δ (ppm): 2.47 (t, 2H), 1.28 (br, 8H), 0.89 (t, 3H).



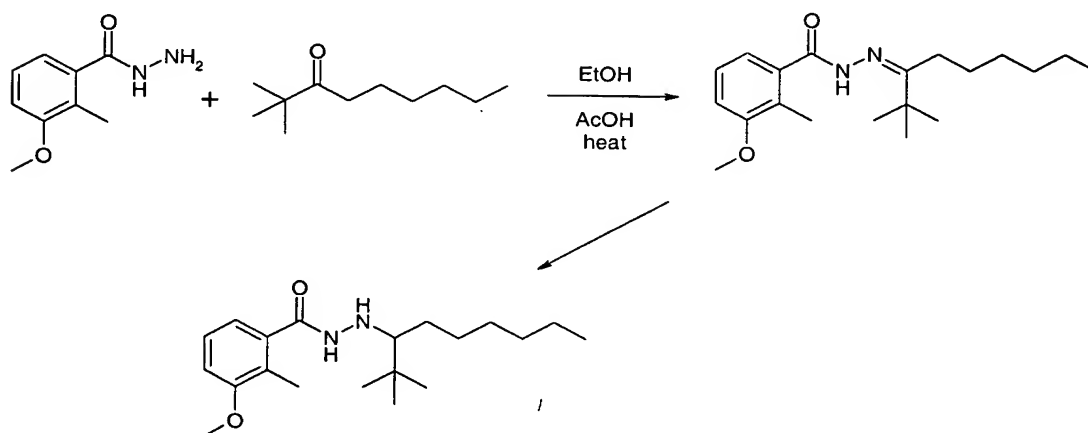
[00257] To a 500 mL, 3-neck flask equipped with magnetic stirring, and chilled in an ice water bath, were added 50 mL CH_2Cl_2 , followed by a solution of 32 g (231 mmol) K_2CO_3 dissolved in 80 mL water, and 24 g (479 mmol) hydrazine hydrate. Over a period of 30-60 minutes, a solution of 20 g (108.3 mmol) 2-methyl, 3-methoxybenzoyl chloride dissolved in 100 mL CH_2Cl_2 were added, while keeping the temperature below 10 °C. The reaction mixture was stirred for an additional hour, during which time a precipitate formed. The precipitate was collected and shaken with a $\text{CH}_2\text{Cl}_2/\text{CHCl}_3$ mixture; the liquid phase was separated from remaining solids, and solvent was removed in vacuo to leave 5.85 g crude product hydrazide. Meanwhile, the original CH_2Cl_2 solution was transferred to a separatory funnel, diluted with CHCl_3 , and shaken with water. The organic layer was removed, washed again with water, dried, and solvent was evaporated to leave a solid residue. This was washed thoroughly with hexane and filtered to provide an additional 6.05 g crude product hydrazide. The combined crude hydrazide was recrystallized from hot ether or ethyl acetate/hexane mixtures to yield 10.04 g 3-methoxy-2-methyl-benzoic acid hydrazide: ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.2 (t, 1H), 6.95 (br s, 1H), 6.9 (m, 2H), 4.15 (br s, 2H), 3.84 (s, 3H), 2.27 (s, 3H).



[00258] Tert-butylcarbazate (80.0 g, 605mmol) was stirred in 800 mL methylene chloride and cooled to 0 °C. To this was added the potassium carbonate solution (937 g, 847 mmol). 2-Ethyl-3-methoxybenzoyl chloride (132 g, 666 mmol) was dissolved in 400 mL of methylene chloride and added to the reaction mixture dropwise over 15 minutes. The mixture was stirred at 0 °C for 15 min, then at room temperature for 18 hours. The reaction was diluted with more methylene chloride and water. The aqueous phase was separated. The remaining organic phase was washed with 1N HCl, water, saturated sodium chloride, dried over magnesium sulfate and evaporated. The residue was triturated with hexane to give a white solid, N'-(2-ethyl-3-methoxy-benzoyl)-hydrazinecarboxylic acid tert-butyl ester (167 g, 567 mmol) in 94% yield. ^1H -NMR (300 MHz, CDCl_3) δ (ppm): 7.4 (br s, 1H), 7.22 (t, 1H), 7.03 (br d, 1H), 6.95 (d, 1H), 6.65 (br, 1H), 3.84 (s, 3H), 2.8 (q, 2H), 1.51 (s, 9H), 1.2 (t, 3H).

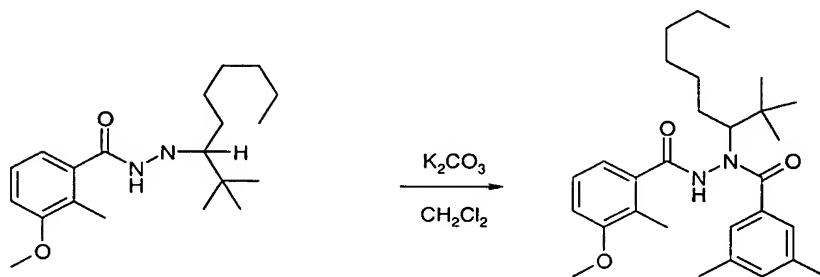


[00259] A round bottom flask was prepared with an overhead stirrer (necessary because of the large amount of solid that precipitates out during the reaction) and a nitrogen inlet. In this flask was stirred N'-(2-ethyl-3-methoxybenzoyl)-hydrazinecarboxylic acid tert-butyl ester (166 g, 564 mmol) in methylene chloride (2260 mL). To this mixture was added trifluoroacetic acid (217 mL, 2820 mmol). The reaction mixture was stirred at room temperature for 18 hours. Ether (1000 mL) and hexane (1000 mL) were added and the mixture was stirred for 1 hour. The precipitate was filtered off and washed with 50% ether/hexane to give a white solid (90.2 g) as the trifluoroacetate salt of the product. The mother liquors and washes were combined, evaporated, and triturated with ether to give an additional amount of a white solid (15.5 g), again as the trifluoroacetate salt of 2-ethyl-3-methoxybenzoic acid hydrazide. The combined yield was 60% (105.7 g). ¹H-NMR (300 MHz, DMSO-d₆) δ (ppm): 7.3 (t, 1H), 7.17 (d, 1H), 6.95 (d, 1H), 3.85 (s, 3H), 2.65 q, 2H), 1.1 (t, 3H). ¹⁹F-NMR (282.4 MHz, DMSO-d₆) δ (ppm): -74.2 (s). Analysis of the free base 2-ethyl-3-methoxybenzoic acid hydrazide: ¹H-NMR (300 MHz, DMSO-d₆) δ (ppm): 9.4 (br s, 1H), 7.2 (t, 1H), 7.05 (d, 1H), 6.85 (d, 1H), 4.45 (br, 2H), 3.85 (s, 3H), 2.6 (q, 2H), 1.1 (t, 3H).



[00260] 3.40g (20 mmol) of 2,2-dimethylnonan-3-one were dissolved in 40 mL of 100% ethyl alcohol. Then 3.60 g (20 mmol) of 2-methyl, 3-methoxybenzoylhydrazide and 20 drops of glacial

acetic acid were added. The reaction mixture was refluxed for 10 hours (required for complete reaction) and monitored by TLC. To a solution of the intermediate 3-methoxy-2-methyl-benzoic acid (1-tert-butyl-heptylidene)-hydrazide, were added 3.5 mL glacial acetic acid and 1.89 g (30 mmol) of sodium cyanoborohydride. The reaction was stirred at room temperature for 24 hours and then refluxed for one hour. The reaction was cooled and 50 mL of water and 10% aqueous NaOH was added until the reaction was basic (pH = ca. 14). Most of the alcohol was removed on a rotary evaporator and the product was extracted with CHCl_3 . The aqueous extract was dried and concentrated to constant weight, yielding 6 g of a viscous material. TLC (1:1 ethyl acetate:hexane) indicated that the material consisted of equal amounts of product alkylated hydrazide, $R_f=0.6$ (1:1 ethyl acetate:hexane) and starting unalkylated hydrazide, $R_f=0.10$ (1:1 ethyl acetate:hexane). Pure product was obtained by column chromatography on silica, 3-methoxy-2-methyl-benzoic acid N'-(1-tert-butyl-heptyl)-hydrazide being eluted with 20-30% ethyl acetate in hexane. ^1H NMR (500 MHz, CDCl_3) δ (ppm): 7.18 (t, 1H), 7.05 (br s, 1H [NH]), 6.91 (d, 1H), 4.9 (br s, NH), 3.84 (s, 3H), 2.5 (m, 1H), 2.28 (s, 3H), 1.2-1.8 (multiple amorphous peaks, 10H), 0.97 (s, 9H), 0.9 (t, 3H).

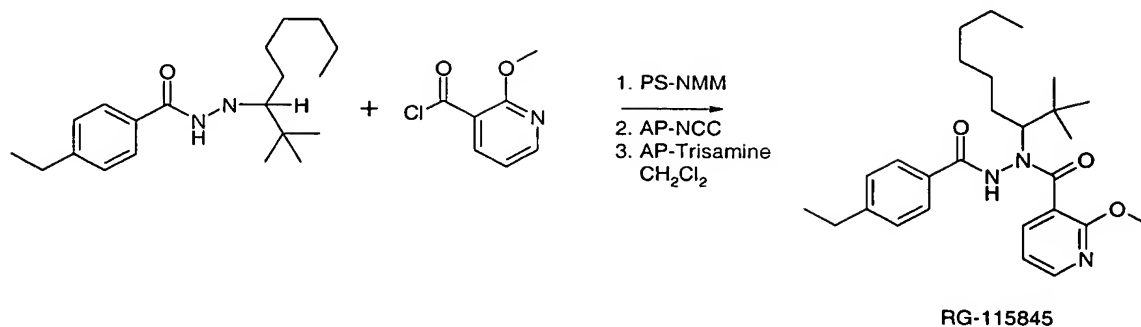


RG-115853

[00261] 164 mg (0.49 mmol) of 3-methoxy-2-methyl-benzoic acid N'-(1-tert-butyl-heptyl)-hydrazide and 82 mg 3,5-dimethylbenzoyl chloride were dissolved in 10 mL CH_2Cl_2 . 7 mL of 25% K_2CO_3 were added, and the reaction mixture was stirred at room temperature overnight. The reaction was monitored by TLC. The phases were separated, adding additional CH_2Cl_2 and/or water as needed to aid manipulation. The CH_2Cl_2 layer was dried and solvent was removed in vacuo to provide 240 mg of crude product. This material was purified by silica gel column chromatography, eluting with a step gradient of 5-20% ethyl acetate in hexane. The desired product eluted in the 15% ethyl acetate fraction, TLC $R_f=0.62$ (1:1 ethyl acetate:hexane), to yield 195 mg 3,5-dimethyl-benzoic acid N-(1-tert-butyl-heptyl)-N'-(3-methoxy-2-methyl-benzoyl)-hydrazide. ^1H NMR (500 MHz, CDCl_3) δ (ppm): 7.2 (s, NH), 7.05 (s, 2H), 7.03 (m, 1H), 7.01 (s, 1H), 6.83 (d, 1H), 6.25 (d, 1H), 4.67 (d, 1H),

3.81 (s, 3H), 2.28 (s, 6H), 1.79 (s, 3H), 1.55 (br m, 2H), 1.3 (br, 8H), 1.34 + 1.07 (s+s, 9H), 0.089 (br s, 3H).

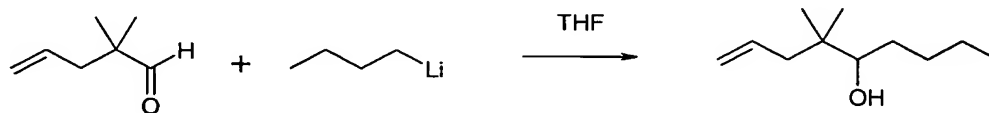
1.2 Preparation of RG-115845



[00262] In a 25 mL vial, 123.1 mg (0.388 mmol) of 4-ethyl-benzoic acid N'-(1-tert-butyl-heptyl)-hydrazide were dissolved in 5 mL CH_2Cl_2 . 3 equivalents of PS-NMM (polystyrene- $\text{SO}_2\text{NH}(\text{CH}_2)_3$ -morpholine, Argonaut Tech.) were added, followed by 3 mL CH_2Cl_2 to create a stirable suspension. One equivalent of acid chloride was added, and the reaction mixture was stirred overnight. The next day, 2 equivalents of AP-NCO resin (isocyanate scavenger, ArgonautTech.) and 2 equivalents of AP-trisamine (polystyrene- $\text{CH}_2\text{NHCH}_2\text{CH}_2\text{NH}(\text{CH}_2\text{CH}_2\text{NH}_2)_2$, Argonaut Tech.) were added with 3 mL CH_2Cl_2 . The mixture was stirred for four hours and the resins were filtered. The reaction was analyzed by TLC, the solvent was removed, and the residue was dried under vacuum. After silica gel column chromatography using a 0-100% ether in hexane gradient, 107.5 mg of 2-methoxy-nicotinic acid N-(1-tert-butyl-heptyl)-N'-(4-ethyl-benzoyl)-hydrazide were isolated.

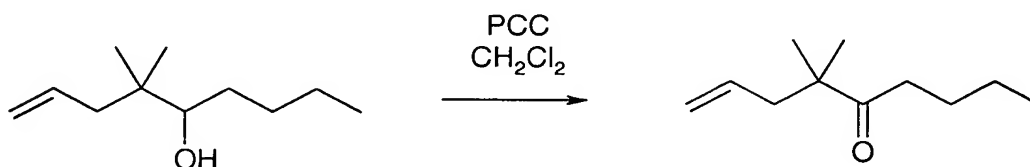
[00263] RG-115843, RG-115844, and RG-115854 were prepared in an analogous manner from corresponding starting hydrazides and acid chlorides.

1.3 Preparation of RG-115878

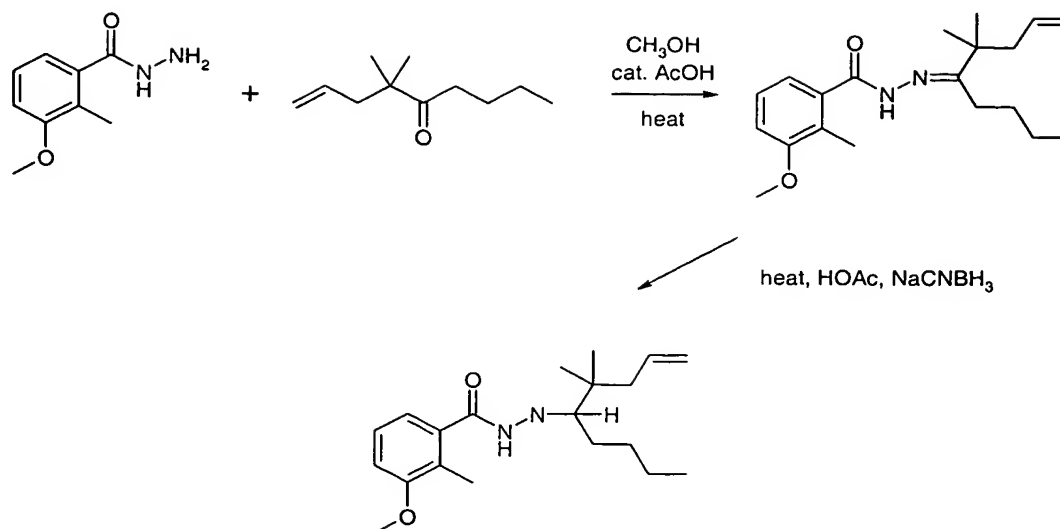


[00264] 15 g (133 mmol) of 2,2-dimethyl-4-pentenal was added to 600 mL of THF in a 2 L, 3-neck flask, flushed with N_2 , and sealed with rubber stoppers. The reaction mixture was cooled to -60°C in

a dry ice/acetone bath. Butyllithium solution (2.5 M in hexane, 64.4 mL, 160.7 mmol) was added 4-5 mL at a time from a 20 mL glass syringe. The reaction temperature was kept at $-65\text{ }^{\circ}\text{C}$, during addition and afterwards was stirred for an additional hour. The reaction mixture was allowed to warm to $-5\text{ }^{\circ}\text{C}$ over 1 hour, cooled to $-60\text{ }^{\circ}\text{C}$ again, and slowly quenched with ammonium chloride solution (10.5 g/200 mL water, 200 mmol). THF was evaporated using a rotary evaporator, keeping the water bath temperature set at $30\text{ }^{\circ}\text{C}$. The resultant aqueous solution was first extracted with CH_2Cl_2 (3 x 200 mL), and then with ether (1 x 200 mL). The organic extracts were combined, dried, and concentrated on a rotary evaporator ($30\text{ }^{\circ}\text{C}$ water bath) to yield 22.64 g of 4,4-dimethyl-non-1-en-5-ol, TLC Rf = 0.67 (1:1 ethyl acetate: hexanes, visualization by I_2). ^1H NMR (CDCl_3 , 500 MHz) δ (ppm): 5.86 (m, 1H), 5.06 (m, 1H), 5.03 (s, 1H), 3.26 (d, 1H), 2.15 (m, 1H), 1.95 (m, 1H), 1.5 (m, 2H), 1.3 (m, 4H), 0.95 (t, 3H), 0.90 (s, 3H), 0.89 (s, 3H).

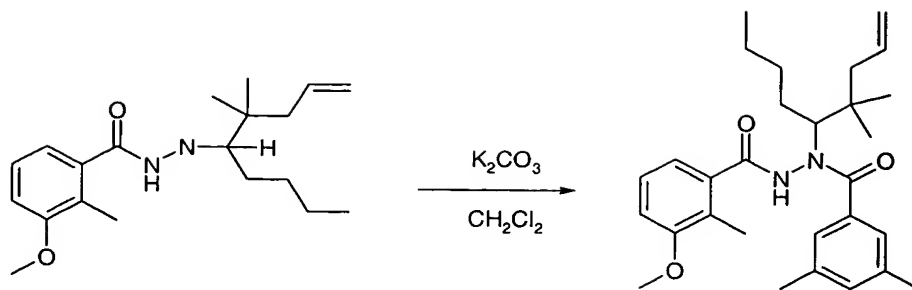


[00265] 17.50 g (102.9 mmol) of 4,4-dimethyl-non-1-en-5-ol and 300 mL of CH_2Cl_2 were added to a 500 mL flask. The reaction was cooled in ice water. While stirring vigorously, 33.29 g (154 mmol) of pyridinium chlorochromate was added in portions (ice bath). The reaction was stirred at room temperature for 24 hours, during which time, it turned black and a sludge formed at the bottom of the flask. Black-brown CH_2Cl_2 was decanted and the sludge was washed twice with CH_2Cl_2 . The reaction product was purified by column chromatography on silica. The brown CH_2Cl_2 mixture was passed through a dry silica column and clean product eluted as the CH_2Cl_2 solution, which after solvent evaporation yielded 12.69 g of product 4,4-dimethyl-non-1-en-5-one. Elution with 5% ethyl acetate /hexane yielded an additional 1.40 g (81% yield). TLC of the pure product gave an Rf = 0.71 (1:1 ethyl acetate:hexane, visualized by I_2). ^1H NMR (CDCl_3 , 500 MHz) δ (ppm): 5.7 (m, 1H), 5.05 (m, 2H), 2.45 (t, 2H), 2.27 (d, 2H), 1.5 (m, 2H), 1.3 (m, 2H), 1.12 (s, 3H), 0.90 (t, 3H).



[00266] A 50 mL round bottom flask was charged with 10.62 g (59 mmol) of 2-methyl-3-methoxybenzoyl hydrazide and 10.0 g (59.5 mmol) of 4,4-dimethylnonan-1-ene, 5-one and 150 mL of methanol. Twenty drops of glacial acetic acid were added as a catalyst and the reaction mixture was refluxed for 9 hours and stirred at room temperature for 48 hours. The intermediate hydrazone was not isolated, but TLC indicated that 30% product was obtained ($R_f = 0.58$, 1:1 hex:ethyl acetate). To the reaction mixture was added 4 mL of acetic acid and 4.72 g (75 mmol) of NaCNBH₃, and the reaction was refluxed for one hour. The reaction mixture was transferred to a 600 mL beaker, and 100 mL water were added, followed by 10% NaOH until basic. Most of the alcohol was evaporated off, and the remaining mixture was extracted with ethyl acetate to yield 12.2 g of residue.

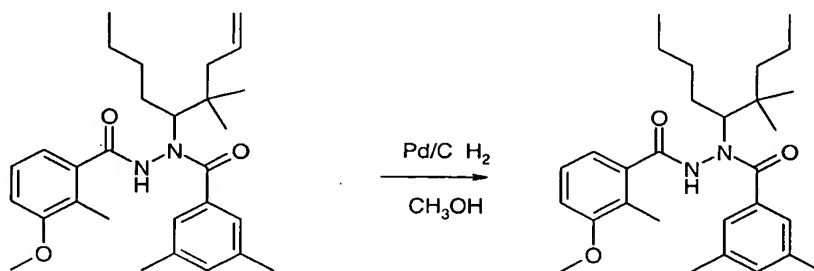
Chromatography on silica gel, eluting with 15-25% ethyl acetate/hexane yielded the crude product hydrazide (2.86 g, 15% yield), two spots by TLC, $R_f = 0.47$ and 0.55 (1:1 ethyl acetate/hexane). Rechromatography yielded pure product by elution with an ethyl acetate/hexane gradient; the 10% ethyl acetate/hexane fraction gave purified 3-methoxy-2-methyl-benzoic acid N'-(1-butyl-2,2-dimethyl-pent-4-enyl)-hydrazide, $R_f = 0.56$, (1:1 ethyl acetate/hexane). ¹H NMR (500 MHz, CDCl₃) δ (ppm): 7.18 (t, 1H), 7.03 (s, 1H, NH), 6.90 (d, 1H), 5.87 (m, 1H), 5.04 (m, 2H), 4.9 (br s, 1H NH), 3.84 (s, 3H), 2.55 (m, 1H), 2.28 (s, 3H), 2.16 (m, 1H), 2.06 (m, 1H), 1.7 (m, 1H), 1.6 (m, 1H), 1.45 (m, 1H), 1.3 (m, 3H), 0.96 (s, 3H), 0.92 (s, 3H), 0.92 (t, 3H).



RG-115878

[00267] 1.8 g (5.42 mmol) of 3-methoxy-2-methyl-benzoic acid N'-(1-butyl-2,2-dimethyl-pent-4-enyl)-hydrazide and 1.34 g (8 mmol) 3,5-dimethylbenzoyl chloride were dissolved in 50 mL CH_2Cl_2 . 20 mL of 20% K_2CO_3 (15 mmol) were added, and the reaction mixture was stirred at room temperature overnight. The phases were separated, adding additional CH_2Cl_2 and/or water as needed to aid manipulation. The CH_2Cl_2 layer was dried and solvent was removed in vacuo to provide 1.65 g of a glassy solid. This material was purified by silica gel column chromatography, eluting with a step gradient of 5-25% ethyl acetate in hexane. The desired product eluted in the 12% ethyl acetate fraction, TLC $R_f=0.59$ (1:1 ethyl acetate:hexane), yield = 1.0 g. Rechromatography again with an ethyl acetate/hexane step gradient provided a purer specimen of the intended diacylhydrazide, 3,5-dimethyl-benzoic acid N-(1-butyl-2,2-dimethyl-pent-4-enyl)-N'-(3-methoxy-2-methyl-benzoyl)-hydrazide, as a white solid. ^1H NMR (500 MHz, CDCl_3) δ (ppm): 7.2 (br s, NH), 7.11 (m, 1H), 7.1 (s, 2H), 7.02 (s, 1H), 6.87 (d, 1H), 6.3 (d, 1H), 5.92 (m, 1H), 5.1 (m, 2H), 4.77 (m, 1H), 3.78 (s, 3H), 2.35 (d, 1H), 2.28 (s, 6H), 2.15 (d, 1H), 1.77 (s 3H), 1.2-1.6 (br m, 6H), 1.05 (s, 3H), 1.03 (s, 3H), 0.94 (t, 3H).

1.4 Preparation of RG-115877



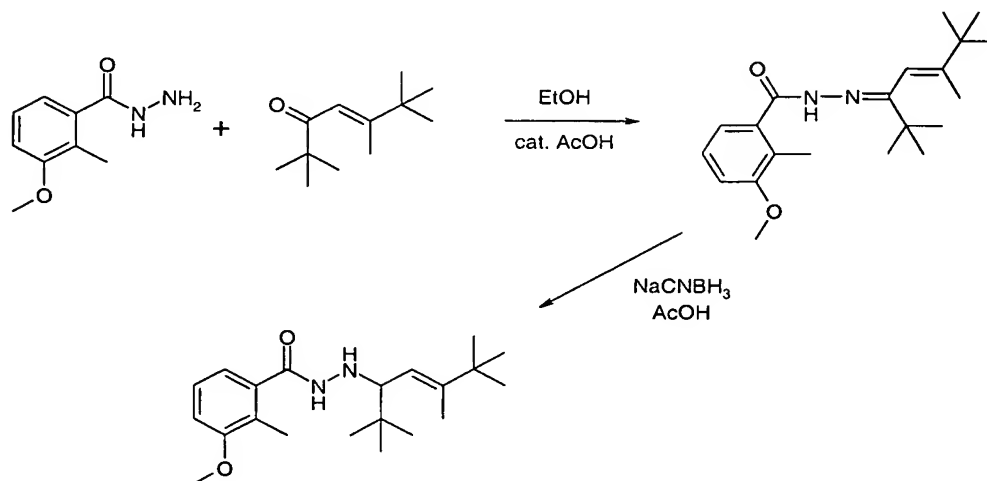
RG-115877

[00268] RG-115878 (240 mg) was hydrogenated in a Parr shaker in methanol under an atmosphere of H_2 using palladium on charcoal catalyst. 3,5-Dimethyl-benzoic acid N-(1-butyl-2,2-dimethyl-

pentyl)-N'-(3-methoxy-2-methyl-benzoyl)-hydrazide was isolated as a white solid: ^1H NMR (500 MHz, CDCl_3) δ (ppm): 7.07 (m, 3H [NH]), 7.00 (m, 2H), 6.83 (d, 1H), 6.29 (d, 1H), 4.73 (d, 1H), 3.79 (s, 3H), 2.29 (s, 6H), 1.78 (s, 3H), 1.58 (br, 4H), 1.38 (br, 6H), 1.03 (m, 6H), 0.96 (m, 6H).

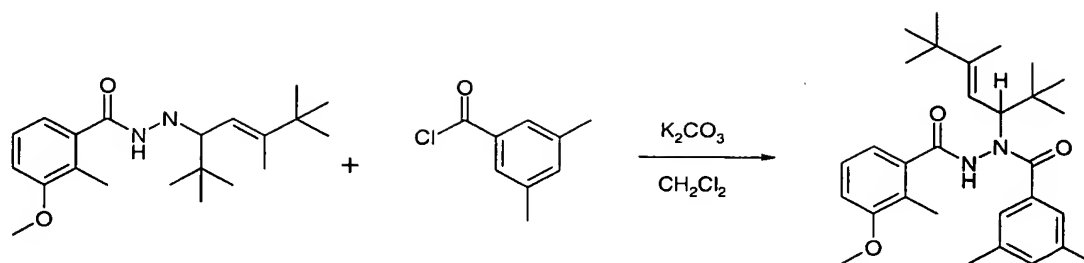
1.5

Preparation of RG-115855



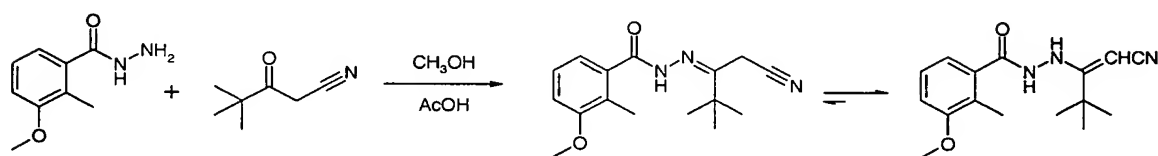
[00269] 1.84 g (100 mmol) of 2,2,5,6,6-pentamethylheptene-3-one (Lancaster Synthesis) was weighed into a 200 mL round bottom flask. 1.80 (10 mmol) of 2-methyl-3-methoxybenzoylhydrazine, 50 mL of ethanol and 20 drops of glacial acetic acid were added. The reaction was refluxed with stirring for 24 hours. The product hydrazone was not isolated but subjected directly to reduction.

[00270] To the 3-methoxy-2-methyl-benzoic acid N'-(1-tert-butyl-3,4,4-trimethyl-pent-2-enyl)-hydrazide reaction mixture was added 3 mL of glacial acetic acid and 950 mg (14.5 mmol) of 95% sodium cyanoborohydride, and the reaction was stirred overnight at room temperature. Most of the ethanol was evaporated, and 50 mL of water were added, and the mixture was basified to pH 14 with 10% NaOH. The remainder of the ethanol was evaporated, and the product was extracted with chloroform. TLC of the residue indicated the presence of the intended product hydrazone ($R_f=0.55$, 1:1 ethyl acetate:hexane), starting hydrazide ($R_f=0.08$), and several minor products. Pure hydrazide was obtained after gradient (ethyl acetate/hexane) silica gel chromatography; the product eluted with 20% ethyl acetate in hexane. Concentration in vacuo yielded 515 mg of white crystalline 3-methoxy-2-methyl-benzoic acid N'-(1-tert-butyl-3,4,4-trimethyl-pent-2-enyl)-hydrazide (15% yield). ^1H NMR (500 MHz, CDCl_3) δ (ppm): 7.2 (t, 1H), 6.93 (s, 1H [NH]), 6.88 (d, 1H), 5.24 (d, 1H), 5.2 (br s, 1H), 3.83 (s, 3H), 3.63 (d, 1H), 2.26 (s, 3H), 1.69 (s, 3H), 1.07 (s, 9H), 0.99 (s, 9H).

**RG-115855**

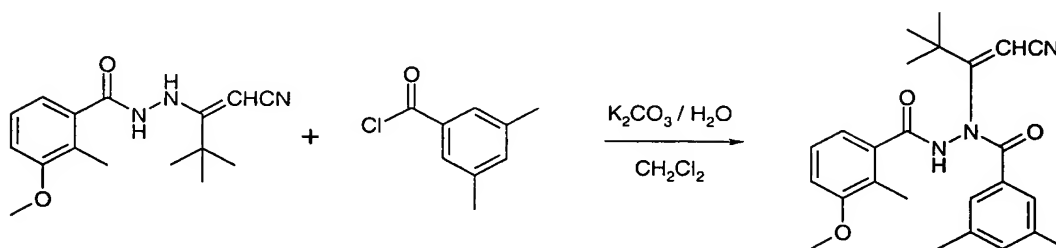
[00271] 219 mg (0.63 mmol) of 3-methoxy-2-methylbenzoic acid N'-(1-tert-butyl-3,4,4-trimethylpent-2-enyl)-hydrazide, 230 mg (1.36 mmol) of 3,5-dimethylbenzoyl chloride, 7 mL of an aqueous 25% K_2CO_3 solution, and 7 mL of CH_2Cl_2 were added to a 20 mL vial. The reaction was stirred at room temperature for 24 hours. The mixture was transferred to a separatory funnel with CH_2Cl_2 , and 45 mL of 17% aqueous K_2CO_3 were added. The CH_2Cl_2 layer was separated, the aqueous layer extracted with 2 x 50 mL portions of CH_2Cl_2 . The organic layers were combined, dried, and concentrated to dryness in vacuo to yield 0.53 g residue, $R_f = 0.67$ (1:1 ethyl acetate: hexane). The residue was chromatographed on silica gel using an ethyl acetate/hexane gradient and the desired product was eluted with 10-11% ethyl acetate in hexane to yield 0.21 g of pure 3-methoxy-2-methylbenzoic acid N'-(1-tert-butyl-3,4,4-trimethylpent-2-enyl)-N'-(3,5-dimethylbenzoyl)-hydrazide. 1H NMR (500 MHz, $CDCl_3$) δ (ppm): 7.2 (1H, [NH]), 7.12 (m, 1H), 7.1 (br s, 2H), 6.95 (s, 1H), 6.85 (d, 1H), 6.6 (d, 1H), 5.55 (d, 1H), 5.45 (d, 1H), 3.8 (s, 3H), 2.3 (s, 6H), 1.95 (s, 3H), 1.6 (s, 3H), 1.07 (s, 9H), 1.02 (s, 9H).

1.6 Preparation of RG-115860



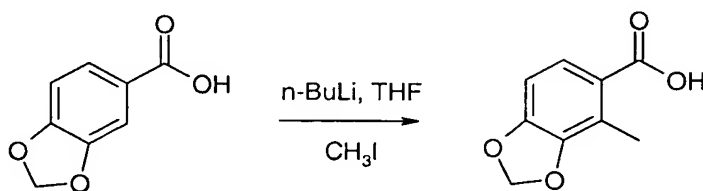
[00272] Tert-butyl-cyanomethylketone (1.75 g, 14 mmol, Lancaster Synthesis) and 2-methyl-3-methoxybenzoylhydrazide (2.52 g, 14 mmol) were added to 20 mL of methanol acidified with 20 drops of glacial acetic acid. The reaction mixture was stirred at room temperature overnight. Solvent was removed in vacuo, the residue was redissolved in CH_2Cl_2 , and the resultant mixture was extracted with aqueous K_2CO_3 . The organic layer was dried and solvent was removed in vacuo. Column chromatography on silica gel, eluting with a 10-35% ethyl acetate/hexane gradient, provided 0.69 g of purified 3-methoxy-2-methylbenzoic acid N'-(1-tert-butyl-2-cyano-vinyl)-hydrazide, as well as a

comparable quantity of somewhat less pure material, TLC: $R_f=0.56$ (1:1 ethyl acetate:hexane). ^1H NMR (500 MHz, CDCl_3) δ (ppm): 7.18 (t, 1H), 7.07 (d, 1H), 6.91 (d, 1H), 5.56 (br s, 2H [NH]), 5.27 (s, 1H), 3.82 (s, 3H), 2.13 (s, 3H), 1.145 (s, 9H).



[00273] 340 mg (1.18 mmol) of 3-methoxy-2-methyl-benzoic acid N' -(1-tert-butyl-2-cyano-vinyl)-hydrazide, 290 mg (1.7 mmol) of 3, 5-dimethylbenzoyl chloride, 3 mL of an aqueous 25% K_2CO_3 solution, and 5 mL of CH_2Cl_2 were stirred at room temperature for 2 days. The mixture was diluted with water and CH_2Cl_2 . The organic layer was separated, and the aqueous layer extracted with additional CH_2Cl_2 . The organic layers were combined, dried, and concentrated to dryness in vacuo to yield 0.72 g semi-solid, $R_f = 0.69$ (1:1 ethyl acetate: hexane, major impurity at $R_f = 0.63$). The residue was chromatographed on silica gel using a 2-20% ethyl acetate/hexane gradient and the desired product was eluted with 4% ethyl acetate in hexane to yield 106 mg of purified 3-methoxy-2-methyl-benzoic acid N' -(1-tert-butyl-2-cyano-vinyl)- N' -(3,5-dimethylbenzoyl) hydrazide. ^1H NMR (500 MHz, CDCl_3) δ (ppm): 7.57 (s, 2H), 7.25 (t, 1H), 7.2 (s, 1H), 7.1 (d, 1H [C=CH]), 7.02 (s, 1H), 6.99 (d, 1H), 3.87 (s, 3H), 2.4 (s, 6H), 2.16 (s, 3H), 1.24 (s, 9H).

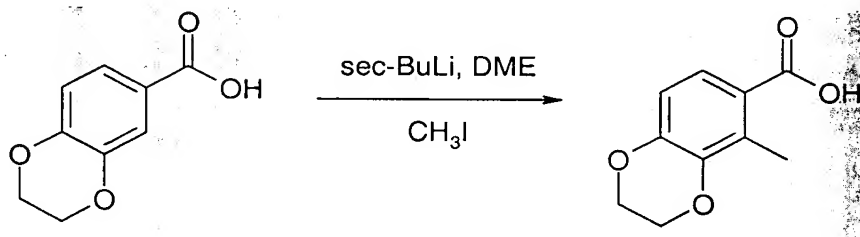
1.7 Preparation of RG-115790



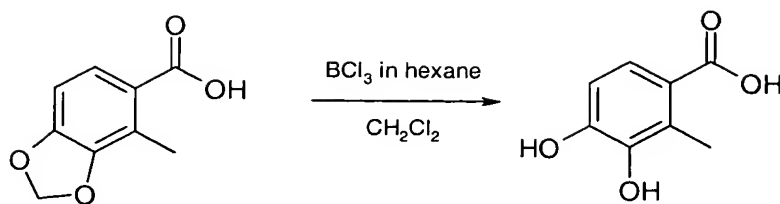
[00274] To a round bottom flask equipped with an overhead stirrer and a nitrogen inlet was added piperonylic acid (50.0g, 301 mmol) and tetrahydrofuran (753 mL). This was cooled to $-75\text{ }^\circ\text{C}$ in a dry ice-acetone bath. N -Butyllithium (1.6 M in hexanes) (414 mL, 662mmol) was added dropwise maintaining the temperature of the reaction at or below $-65\text{ }^\circ\text{C}$. When the addition was complete the cooling bath was removed and the reaction was allowed to warm to $-20\text{ }^\circ\text{C}$. The reaction was returned

to the bath and cooled to at least $-60\text{ }^{\circ}\text{C}$ at which time iodomethane (37.5 mL, 602 mmol) was added dropwise. Cooling was removed and the reaction allowed to warm to $10\text{ }^{\circ}\text{C}$ at which time the reaction was placed in an ice bath and stirred at $0\text{ }^{\circ}\text{C}$ for 30 minutes. The reaction was quenched by addition of 1N HCl and the tetrahydrofuran removed by evaporation. The residue was slurried in 1N HCl and filtered. The filtrate was washed with water, and dried at $50\text{--}60\text{ }^{\circ}\text{C}$ in vacuo to give a pale yellow solid, 4-methyl-benzo[1,3]dioxole-5-carboxylic acid (52.8 g, 293 mmol) in 97% yield. $^1\text{H-NMR}$ (300 MHz, CD_3COCD_3) δ (ppm): 2.44 (s, 3H), 6.10 (s, 2H), 6.77 (d, 1H), 7.64 (d, 1H). TLC R_f = 0.54 (1:1 ethyl acetate/hexane, piperonylic acid = 0.41).

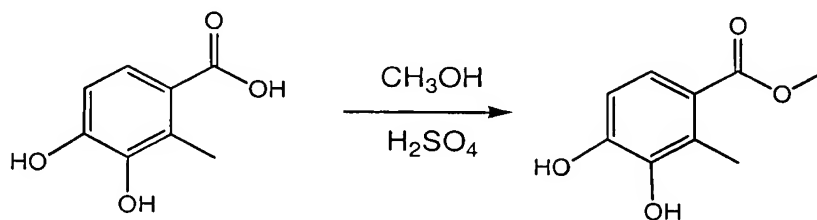
[00275] Notes: Overhead stirring, as opposed to magnetic, is necessary as the carboxylate salt forms a heavy precipitate. Addition of the first equivalent of butyllithium is rather exothermic (deprotonation of the carboxylic acid). The rate of butyllithium addition can be substantially increased for the second equivalent. Reaction temperatures of $-60\text{ }^{\circ}\text{C}$ or higher prior to completion of the butyllithium addition result in formation of butylketone in increasing amounts. The addition of iodomethane is exothermic, with a kick off temperature of approximately -10 to $0\text{ }^{\circ}\text{C}$ at which point a temperature increase of $10\text{--}15\text{ }^{\circ}\text{C}$ occurs.



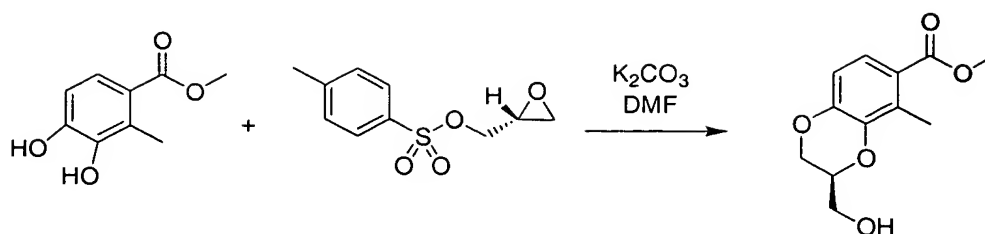
[00276] To a round bottom flask equipped with magnetic stirring, an addition funnel and a nitrogen inlet, was added benzo(1,4)dioxan-6-carboxylic acid (18.00g, 99.91 mmol) and 1,2-dimethoxyethane (667 mL). This mixture was cooled to $-75\text{ }^{\circ}\text{C}$ in a dry ice-acetone bath. To this was added 1.3 M sec-butyl lithium in cyclohexane (230.6 mL, 299.7 mmol) over 1 hour, maintaining reaction temperature below $-60\text{ }^{\circ}\text{C}$. The reaction was removed from the cooling bath, allowed to warm to $-20\text{ }^{\circ}\text{C}$, and subsequently stirred at $-20\text{ }^{\circ}\text{C}$ for 45 min. The reaction was cooled to $-50\text{ }^{\circ}\text{C}$, and iodomethane (15.6 mL, 249.8 mmol) was added. The reaction was again removed from the cooling bath, allowed to warm to $-20\text{ }^{\circ}\text{C}$, and stirred at this temperature for 45 min. All cooling was removed and the reaction stirred at room temperature for 16 hours. The reaction was quenched by addition of a few mls of 1N HCl (aq) and the solvent removed by evaporation. The residue was made substantially acidic by the addition of aqueous 1N HCl. The resultant precipitate was filtered and washed with water to give a light brown solid, 5-methyl-benzo(1,4)dioxan-6-carboxylic acid, (6.60 g, 33.9 mmol) in 34% yield. $^1\text{H-NMR}$ (300 MHz, CDCl_3) δ (ppm): 7.62 (d, 1H), 6.8 (d, 1H), 4.30 (br s, 4H), 2.52 (s, 3H).



[00277] To a round bottom flask equipped with magnetic stirring, a 25% NaOH gas trap, and a nitrogen inlet was added 2-methylpiperonylic acid (26.0 g, 144 mmol) and methylene chloride (482 mL). To this was added the boron trichloride solution. The mixture was stirred at room temperature for 3 hours. The reaction was quenched by the careful addition of 500 mL of water. **CAUTION:** this step causes foaming and generates large amounts of HCl. The reaction was stirred for 30 min. The layers were separated, the TLC of the organic layer was checked to confirm the absence of product and the organic layer was subsequently discarded. The aqueous layer was extracted 3 times with ethyl acetate. The combined extracts were washed once with water, once with saturated NaCl solution, dried over sodium sulfate, filtered, and evaporated to give a tan solid, 3,4-dihydroxy-2-methylbenzoic acid, (20.0 g, 119 mmol) in 82% yield. Note: BCl₃ in heptane has been used as well with equal success. ¹H-NMR (300 MHz, CD₃COCD₃) δ (ppm): 2.51 (s, 3H), 6.76 (d, 1H), 7.45 (overlapping s+d, 1H+1H), 9.06 (s, 1H). TLC R_f = 0.20 (1:1 ethyl acetate/hexane).

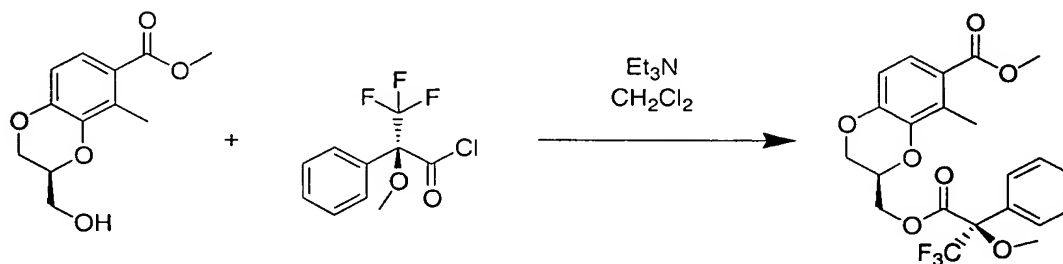


[00278] To a round bottom flask equipped with magnetic stirring and a reflux condenser was added 3,4-dihydroxy-2-methylbenzoic acid (23.3 g, 139 mmol), methanol (139 mL), and concentrated sulfuric acid (0.77 mL, 13.9 mmol). This mixture was heated to reflux for 18 hours. The reaction was cooled to room temperature and the methanol evaporated. The residue was taken up in ether and washed once with saturated sodium bicarbonate solution, once with saturated sodium chloride solution, dried over sodium sulfate, filtered and evaporated to give a brown solid, methyl-3,4-dihydroxy-2-methylbenzoate, (20.9 g, 114 mmol) in 83% yield. ¹H-NMR (300 MHz, CD₃COCD₃) δ (ppm): 2.47 (s, 3H), 3.79 (s, 3H), 6.78 (d, 1H), 7.46 (d, 1H); TLC: R_f = 0.54 (1:1 ethyl acetate/hexane).



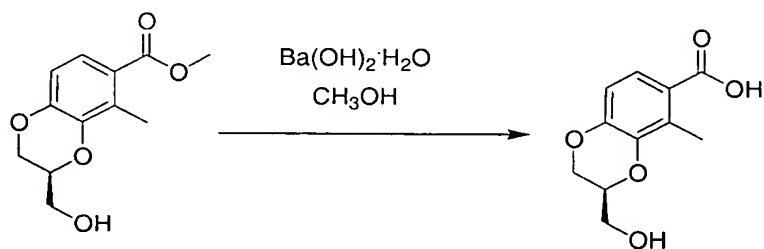
[00279] To a round bottom flask equipped with magnetic stirring and a nitrogen inlet was added methyl 3,4-dihydroxy-2-methylbenzoate (24.0 g, 132 mmol), (2R)-(-)-glycidyl tosylate (30.1 g, 132 mmol), potassium carbonate (21.8 g, 158 mmol), and DMF (264 mL). This mixture was heated to 60 °C for 5 hours. The mixture was cooled to room temperature, diluted with ether, and washed once with water. The aqueous wash was back extracted once with ether and the ethereal solutions combined. These organic solutions were then washed 3 times with saturated sodium chloride solution, dried over magnesium sulfate, filtered and evaporated to give a brown oil. This oil was chromatographed on silica gel eluting with 25% ether in hexanes to give a yellow oil, 3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid methyl ester, (24.6 g, 103 mmol) in 78% yield. ¹H-NMR (300 MHz, CD₃COCD₃) δ (ppm): 2.42 (s, 2H), 3.80 (s, 3H), 3.83 (m, 1H), 4.12 (dd, 1H), 4.23 (m, 2H), 4.42 (dd, 1H), 6.75 (d, 1H), 7.42 (d, 1H); TLC R_f = 0.40 (1:1 ethyl acetate / hexane).

[00280] Notes: (a) (2S)-(+)-Glycidyl tosylate can be used under identical conditions to give the opposite stereochemistry, (b) Formation of a Mosher's ester of this compound indicates the presence of a single regio and stereoisomer by ¹⁹F NMR, (c) X-ray crystal structure determination of the amide formed from (R)-(+)-1-(1-naphthyl)ethylamine confirmed the indicated regio and stereochemistry for the product 3-hydroxymethylbenzodioxan. Delgado, A.; Leclerc, G.; Lobato, C.; Mauleon, D. *Tetrahedron Lett.* **1988**, 29(30), 3671.

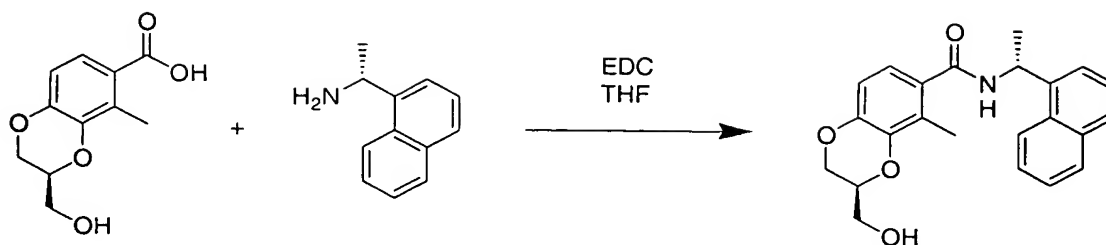


[00281] To a round bottom flask equipped with magnetic stirring and a nitrogen inlet was added 3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid methyl ester (0.10g, 0.44mmol) and dry methylene chloride (1.5 mL). The mixture was cooled in an ice bath and

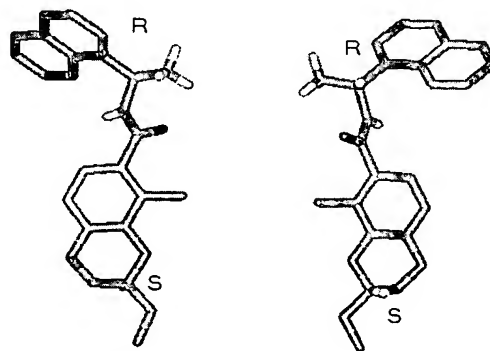
triethylamine (0.12 mL, 0.87 mmol) was added followed by (R)-(-)- α -methoxy- α -trifluoromethylphenylacetyl chloride (0.09 mL, 0.48 mmol). The mixture was stirred for 2 hours at which point TLC showed the reaction to be incomplete. More (R)-(-)- α -methoxy- α -trifluoromethylphenylacetyl chloride (0.09 mL, 0.48 mmol) was added and the reaction was allowed to stir 1 hour at room temperature. The mixture was diluted with methylene chloride, washed once with 1N HCl, once with saturated sodium bicarbonate, dried over magnesium sulfate, filtered and evaporated. The residue was chromatographed on silica gel eluting with methylene chloride to give a clear oil, 5-methyl-3-(3,3,3-trifluoro-2-methoxy-2-phenyl-propionyloxymethyl)-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid methyl ester, for which a yield was not determined. $^1\text{H-NMR}$ (300 MHz, CD_3COCD_3) δ (ppm): 2.38 (s, 3H), 3.58 (s, 3H), 3.81 (s, 3H), 4.10 (m, 1H), 4.48 (dd, 1H), 4.65 (m, 2H), 4.86 (m, 1H) 6.78 (d, 1H) 7.47 (m, 5H), 7.60 (d, 1H); $^{19}\text{F-NMR}$ (300 MHz, CD_3COCD_3) δ (ppm): -72.89 (s); both $^1\text{H-}$ and $^{19}\text{F-NMR}$ indicate the presence of only one stereo- and regioisomer; TLC: R_f = 0.62 (1:1 ethyl acetate/hexane).



[00282] To a round bottom flask equipped with magnetic stirring was added 3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid methyl ester (6.40 g, 26.9 mmol), barium hydroxide monohydrate (25.44 g, 134.3 mmol), and methanol (108 mL). This mixture was stirred at room temperature for 24 hours. The methanol was evaporated. The resulting slurry was taken up in water and washed once with ether. The aqueous layer was acidified by pouring into iced concentrated HCl. This was extracted twice with ethyl acetate. The ethyl acetate extracts were combined, dried over magnesium sulfate, filtered and evaporated to give a yellow solid, 3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid, (4.82 g, 21.5 mmol) in 80% yield. $^1\text{H-NMR}$ (300 MHz, CD_3COCD_3) δ (ppm): 2.45 (s, 3H), 3.85 (m, 2H), 4.12 (m, 1H), 4.24 (m, 1H), 4.41 (m, 1H), 6.76 (d, 1H), 7.51 (d, 1H); TLC: R_f = 0.32 (1:1 ethyl acetate/hexane).



[00283] To a round bottom flask equipped with magnetic stirring and a nitrogen inlet was added 3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid (2.00g, 8.92mmol), THF (45 mL), (R)-(+)-1-(1-naphthyl)ethylamine (2.23 mL, 8.92 mmol), and finally 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC) (1.53 g, 9.81 mmol). This mixture was stirred at room temperature for 36 hours. A few drops of water were added and the mixture stirred for 5 minutes. The reaction was diluted with ether and washed once with 1N HCl, once with saturated sodium bicarbonate(aq), once again with 1N HCl, dried over magnesium sulfate, filtered and evaporated to provide 3-hydroxymethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carboxylic acid (1-naphthalen-1-yl-ethyl)-amide. Crystals of acceptable purity for x-ray crystallography were obtained by subliming the material twice under vacuum at 60 °C. This x-ray analysis established the regiochemistry and the absolute stereochemistry of the compounds as shown, since the configuration of the naphthylethylamine stereocenter is known and not subject to racemization under the conditions of synthesis. ¹H-NMR (300 MHz, CD₃COCD₃) δ (ppm): 1.71 (d, 3H), 2.22 (s, 3H), 3.80 (m, 2H), 4.06 (m, 1H), 4.23 (m, 2H), 4.32 (m, 1H), 6.09 (m, 1H), 6.65 (d, 1H), 6.87 (d, 1H), 7.55 (m, 3H), 7.69 (m, 2H), 7.83 (d, 1H), 7.94 (d, 1H), 8.35 (d, 1H); TLC R_f = 0.12 (1:1 ethyl acetate/hexane).



RG-119097

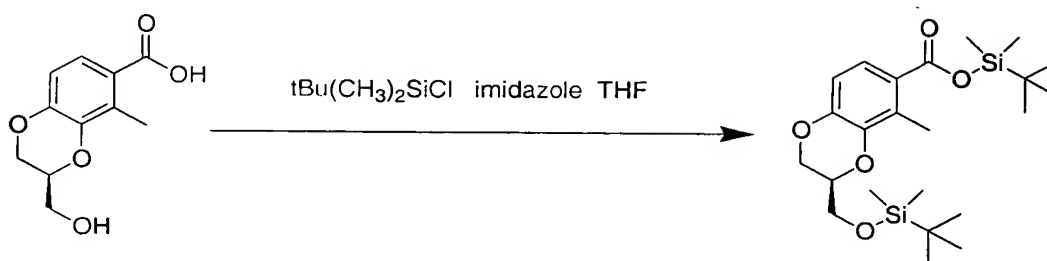
Crystal Data for:	MSC01712 (RG-119097)
Date of Entry:	09-20-2001
Reference:	Unpublished

Synthesis:	RHeoGene
Crystallography by:	J. C. Huffman (huffman@indiana.edu)
Compound Name:	not known
Formula:	C ₂₃ H ₂₃ N ₄ O
Empirical Formula:	C ₂₃ H ₂₃ N ₄ O
Color of Crystal:	Colorless
Crystal System:	Monoclinic
Space Group:	P 21
Cell Dimensions:	(at -161. C; 3421 peaks)
a =	11.9081(18)
b =	4.8854(7)
c =	16.3819(25)
alpha =	90.00(0)
beta =	97.758(4)
gamma =	90.00(0)
Z (Molecules/Cell):	2
Volume:	944.31
Calculated Density:	1.327
Molecular Weight:	377.44
Linear Absorption Coefficient:	0.906
Final Residuals are:	
R(F)=	.062
Rw(F) =	.050

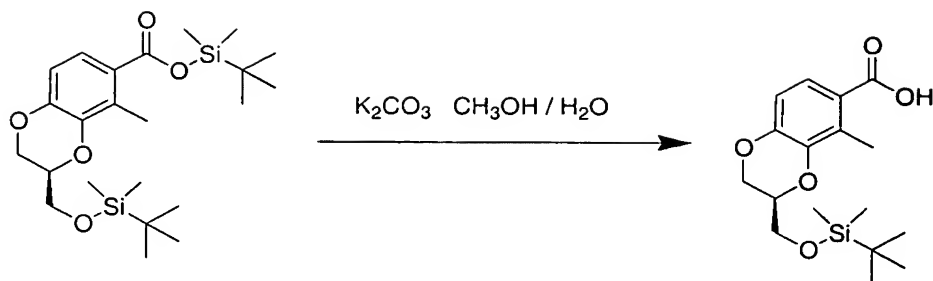
XYZ fractional coordinates:

C(1)	12796(3)	9729*	835(3)	22
C(2)	11822(3)	8046(11)	622(3)	24
C(3)	11171(4)	8270(12)	-122(3)	31
C(4)	11442(4)	10186(11)	-712(3)	33
C(5)	12366(4)	11786(12)	-544(3)	35
C(6)	13084(3)	11594(12)	238(3)	28
C(7)	14062(4)	13263(11)	419(3)	33
C(8)	14737(4)	13001(11)	1149(3)	32
C(9)	14460(3)	11138(11)	1749(3)	27
C(10)	13510(3)	9560(10)	1616(2)	20
C(11)	13203(3)	7583(11)	2258(2)	21
C(12)	14097(3)	7159(13)	2994(3)	30
N(13)	12130(2)	8376(10)	2530(2)	18
C(14)	11390(3)	6516(11)	2737(2)	21
O(15)	11632(2)	4097(9)	2774(2)	34
C(16)	10259(3)	7575(10)	2900(2)	18
C(17)	9704(3)	9513(11)	2367(2)	22
C(18)	8643(3)	10491(11)	2469(2)	23
C(19)	8117(3)	9439(10)	3115(2)	20
O(20)	7068(2)	10502(9)	3208(2)	26
C(21)	6810(3)	10089(11)	4020(3)	25
C(22)	6986(3)	7197(11)	4283(2)	24
C(23)	6706(3)	6660(12)	5139(2)	22
O(24)	5518(2)	6959(9)	5166(2)	26
O(25)	8150(2)	6406(9)	4280(2)	20
C(26)	8647(3)	7491(10)	3634(2)	18
C(27)	9726(3)	6479(11)	3539(2)	18

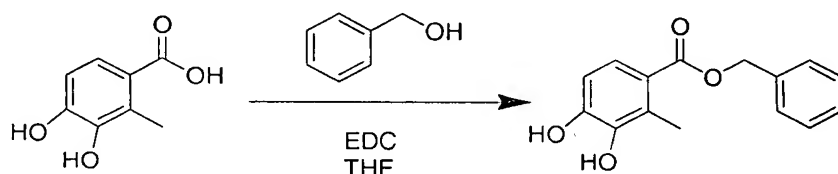
C(28)	10270(4)	4441(12)	4154(3)	22
H(36)	1306(3)	557(7)	201(2)	15(7)
H(45)	641(3)	577(8)	388(2)	33(10)
H(29)	1168(3)	636(9)	96(3)	35(10)
H(30)	1045(3)	727(8)	-22(2)	34(10)
H(31)	1095(3)	1040(8)	-121(3)	27(9)
H(32)	1264(3)	1319(9)	-89(3)	39(10)
H(33)	1417(3)	1460(9)	1(3)	37(10)
H(34)	1547(3)	1404(8)	132(2)	27(9)
H(35)	1499(4)	1049(12)	220(3)	71(16)
H(37)	1488(3)	652(8)	277(3)	32(9)
H(38)	1424(3)	896(9)	326(3)	37(11)
H(39)	1388(3)	543(8)	334(2)	28(9)
H(40)	1187(3)	995(7)	242(2)	5(7)
H(41)	1010(3)	1024(9)	200(3)	34(10)
H(42)	824(3)	1205(8)	209(2)	23(8)
H(43)	738(3)	1156(9)	439(2)	33(9)
H(44)	601(2)	1061(6)	404(2)	5(6)
H(46)	699(4)	820(11)	549(3)	52(12)
H(47)	702(3)	480(8)	538(2)	20(8)
H(48)	509(4)	508(10)	512(3)	41(11)
H(49)	970(5)	414(13)	455(5)	100(21)
H(50)	1010(3)	274(10)	398(3)	32(10)
H(51)	1099(3)	470(8)	420(2)	26(9)



[00284] To a round bottom flask equipped with magnetic stirring and a nitrogen inlet was added 3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid (1.40 g, 6.24 mmol), THF (32 mL), and t-butyldimethylsilyl chloride (5.65 g, 18.73 mmol). This mixture was stirred at room temperature for 30 min. Imidazole (1.87 g, 27.47 mmol) was added and the reaction mixture stirred for another 30 min. The mixture was diluted with hexane and washed once with 1 N HCl, once with brine, dried over magnesium sulfate and evaporated to give a yellow oil. This was purified by flash chromatography (silica gel, 5% ether/hexane) to give a yellow oil, 3-(tert-butyldimethylsilyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid tert-butyldimethylsilyl ester, (1.26 g, 2.78 mmol) in 45% yield. ¹H-NMR (300 MHz, CD₃COCD₃) δ (ppm): 0.10 (d, 6H), 0.33 (s, 6H), 0.90 (s, 9H), 0.99 (s, 9H), 2.45 (s, 3H), 3.95 (m, 2H), 4.13 (m, 1H), 4.26 (m, 1H), 4.40 (dd, 1H), 6.75 (d, 1H), 7.51 (d, 1H); TLC: R_f = 0.87 (1:1 ethyl acetate/hexane).

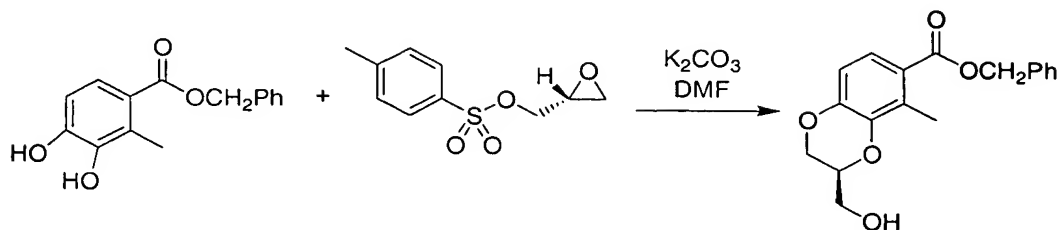


[00285] To a round bottom flask equipped with magnetic stirring and a nitrogen inlet was added the starting material (1.40 g, 3.09 mmol), methanol (16 mL), and 25% potassium carbonate (aqueous) (5 mL). This mixture was stirred at room temperature for 4 hours. The methanol was evaporated and ethyl acetate added to the residue. The mix was acidified with 1N HCl and the layers separated. The aqueous layer was extracted once more with ethyl acetate. The combined ethyl acetate layers were washed once with brine, dried over magnesium sulfate, and evaporated to give a yellow solid, 3-(tert-butyldimethylsilyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid tert-butyldimethylsilyl ester, (0.93 g, 2.75 mmol) in 89% yield. 1H -NMR (300 MHz, CD_3COCD_3) δ (ppm): 0.10 (s, 6H), 0.90 (s, 9H), 1.94 (s, 3H), 3.95 (m, 2H), 4.13 (m, 1H), 4.26 (m, 1H), 4.40 (dd, 1H), 6.75 (d, 1H), 7.51 (d, 1H); TLC R_f = 0.62 (1:1 ethyl acetate/hexane).

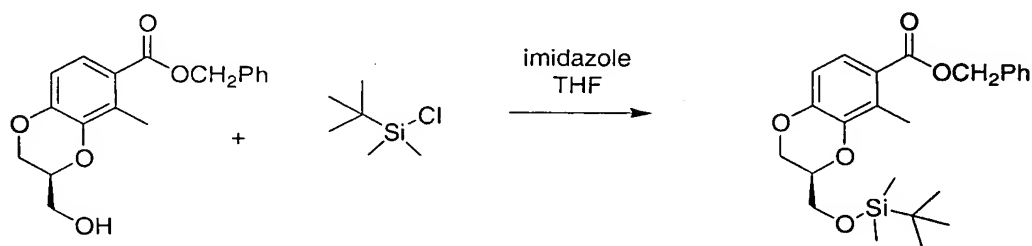


[00286] To a round bottom flask equipped with magnetic stirring and a nitrogen inlet was added 3,4-dihydroxy-2-methylbenzoic acid (23.0 g, 137 mmol), benzyl alcohol (21.2 mL, 205 mmol), and THF (274 mL). To this stirred mixture was added 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC) (30.2 g, 157 mmol). This mixture was stirred at room temperature 36 hours. A few milliliters of water were added and the mixture was stirred for 10 min whereupon the THF was evaporated. The residue was partitioned between ethyl acetate and 1N HCl. The layers were separated and the aqueous layer extracted twice with ethyl acetate. The combined organic extracts were washed once with saturated sodium bicarbonate solution, once with 1N HCl, and once with brine. The organic solution was then dried over sodium sulfate, filtered and evaporated. The resulting oil was triturated with 50% methylene chloride in hexanes from which a tan solid was filtered in 43% yield, benzyl 3,4-dihydroxy-2-methylbenzoate (15.2 g, 59 mmol). 1H -NMR (300

MHz, CD_3COCD_3) δ (ppm): 2.49 (s, 3H), 5.29 (s, 2H), 6.78 (d, 1H), 7.43 (m, 6H); TLC: R_f = 0.49 (1:1 ethyl acetate/hexane).

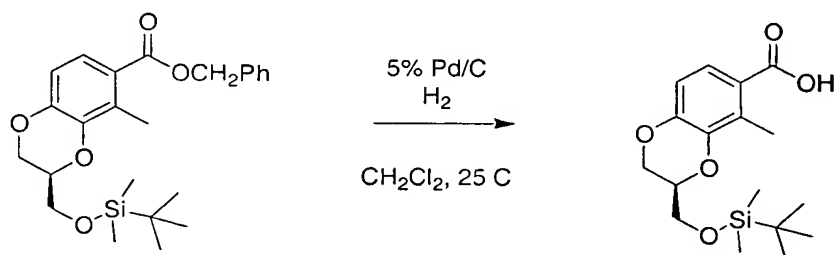


[00287] Into a 500 mL round bottom flask, were added 25.21 g (97.71 mmol) benzyl 3,4-dihydroxy-2-methylbenzoate, 22.28 g (97.71 mmol) of (2R)-(-)-glycidyl tosylate, 16.56 g (120 mmol) K_2CO_3 , and 200 mL DMF. The reaction mixture was stirred at 65 °C in an oil bath for 5 hours. The reaction was allowed to cool, diluted with water (750 mL), and extracted with ether (3 x 300 mL). The ether phase was back-extracted with 100 mL water, and the combined organic phases were dried and evaporated to yield 31.1 g oil. TLC indicated completeness of the reaction: R_f = 0.36 (1:1 ethyl acetate:hexane). Column chromatography using a gradient of 10-50 % ethyl acetate in hexane provided 23.0 g pure 3-hydroxymethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carboxylic acid benzyl ester (80% yield). ^1H NMR (300 MHz, CD_3OD) δ (ppm): 7.4 (m, 6H), 6.75 (d, 1H), 5.28 (s, 2H), 4.4 (d, 1H), 4.2 (m, 1H), 4.1 (dd, 1H), 3.8 (m, 2H), 2.46 (s, 3H).

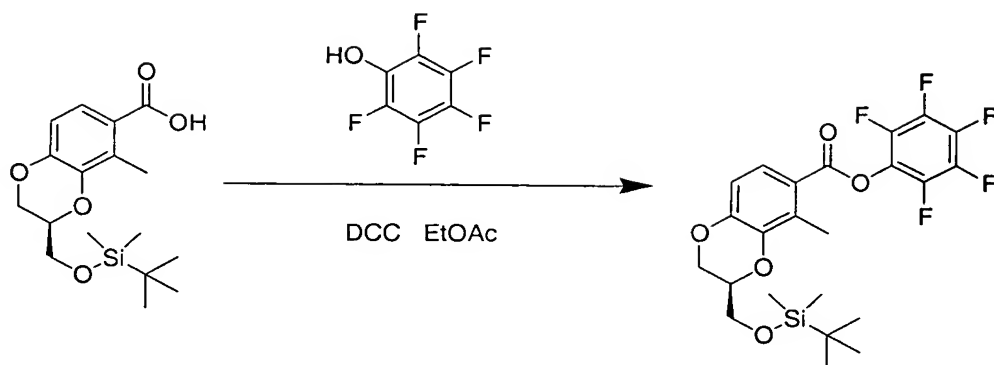


[00288] 23 g (78.2 mmol) of 3-hydroxymethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carboxylic acid benzyl ester was dissolved in 300 mL dry THF. While stirring, 18.08 g t-butyldimethylsilyl chloride and 10.2 g imidazole were added. The mixture was stirred at room temperature overnight, during which time a white precipitate, imidazole-HCl developed. Monitoring by TLC (1:1 ethyl acetate:hexane) demonstrated progress of the reaction. The white precipitate was filtered off, and the remaining THF solution was concentrated on a rotary evaporator. The residue was redissolved in 300 mL CH_2Cl_2 , and extracted with 300 mL water. The CH_2Cl_2 extract was dried and solvent was removed in vacuo to yield 26.55 of the silyl ether product, TLC: R_f = 0.7 (1:1 ethyl

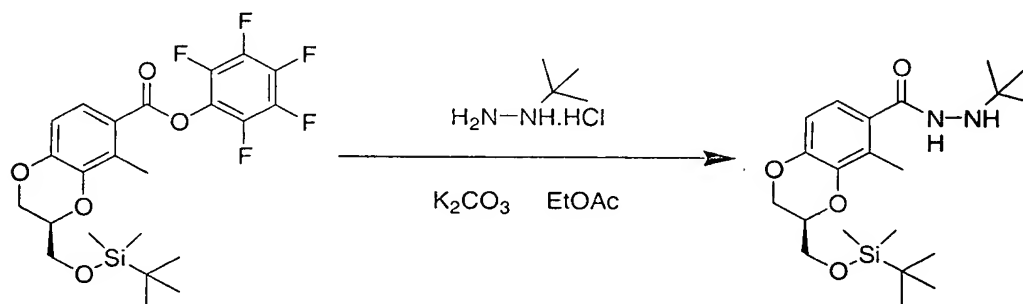
acetate:hexane); crude yield = 83%. The product was redissolved in 150 mL hexane and chromatographed by silica gel column chromatography, eluting with hexane, followed by 2% ethyl acetate in hexane. 21.28 g pure 3-(tert-butyl-dimethyl-silanyloxymethyl)-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carboxylic acid benzyl ester was recovered (67% yield). ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.45 (d, 1H), 7.3 (m, 6H), 6.65 (d, 1H), 4.27 (d, 1H), 4.12 (m, 1H), 4.05 (dd, 1H), 4.85 (dd, 1H), 4.75 (dd, 1H), 2.37 (s, 3H), 0.81 (s, 9H), 0.01 (s, 3H).



[00289] 9.63 g of 3-(tert-Butyl-dimethyl-silanyloxymethyl)-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carboxylic acid benzyl ester was dissolved in ca. 125 mL of dry CH_2Cl_2 . The solution was transferred to a Parr hydrogenation bottle, and 5 g of 5% Pd on carbon were added. The bottle was charged with hydrogen and shaken on a Parr hydrogenation apparatus for 3 hours; hydrogen uptake ceased after 2 hours, as indicated by pressure monitoring. 100 mL CHCl_3 were added, and the Pd/C was filtered off after adding MgSO_4 as an aid to increase particle size. The CH_2Cl_2 - CHCl_3 phase was washed with 200 mL of 0.1N HCl to remove the by-product benzyl alcohol. The organic layer was dried and evaporated to give 6.47 g 3-(tert-butyl-dimethyl-silanyloxymethyl)-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carboxylic acid, after trace solvent removal in a vacuum oven. TLC R_f =0.55 (1:1 ethyl acetate:hexane). Note: (a) Pd catalyzed hydrogenation is a far superior method to triethylsilane/palladium diacetate for this benzyl ester cleavage, (b) washing product with weak aqueous acid is a good way to remove benzyl alcohol. ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.55 (d, 1H), 6.7 (d, 1H), 4.3 (d, 1H), 4.17 (m, 1H), 4.02 (dd, 1H), 3.82 (dd, 1H), 3.75 (dd, 1H), (2.45 (s, 3H), 0.8 (s, 9H), 0.01 (s, 3H).

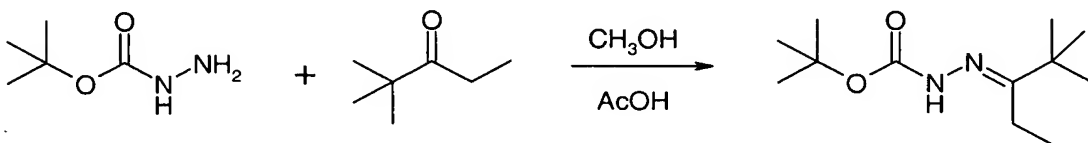


[00290] To a round bottom flask equipped with magnetic stirring and a nitrogen inlet was added 3-(tert-butyldimethylsilyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid (0.72 g, 2.13 mmol) and ethyl acetate (11 mL). To this was added pentafluorophenol (0.82 g, 2.23 mmol) followed by 1 M dicyclohexylcarbodiimide in methylene chloride (2.34 mL, 2.34 mmol). This mixture was stirred at room temperature for four hours. A few milliliters of water were added and stirring continued for 10 min. The reaction was filtered, diluted with ethyl acetate, washed once with 1 N HCl, once with saturated aqueous sodium bicarbonate solution, once with saturated sodium chloride, dried over magnesium sulfate, filtered and evaporated to give an oil. This oil was flash chromatographed on silica gel eluting with hexanes to give a clear colorless oil, 3-(tert-butyldimethylsilyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid pentafluorophenyl ester, (0.90 g, 1.78 mmol) in 84% yield. $^1\text{H-NMR}$ (300 MHz, CDCl_3) δ (ppm): 0.10 (s, 6H), 0.90 (s, 9H), 2.50 (s, 3H), 3.90 (m, 2H), 4.14 (m, 1H), 4.24 (m, 1H), 4.40 (m, 1H), 6.83, (d, 1H), 7.86, (d, 1H); TLC R_f = 0.88 (1:1 ethyl acetate/hexane).

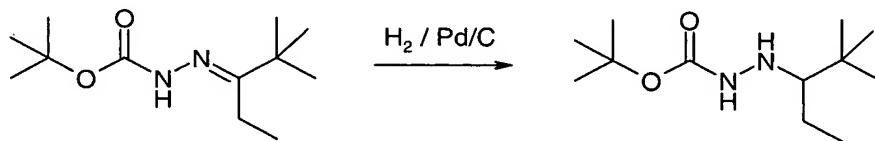


[00291] To a round bottom flask equipped with magnetic stirring was added ethyl acetate (5 mL) and 25 wt% aqueous potassium carbonate solution (2.96 g, 5.35 mmol). To this was added t-butylhydrazine hydrochloride (0.33 g, 2.68 mmol) followed by 3-(tert-butyldimethylsilyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid pentafluorophenyl ester (0.90 g, 1.78 mmol) dissolved in ethyl acetate (4 mL). This mixture was stirred at room temperature for 18 hours.

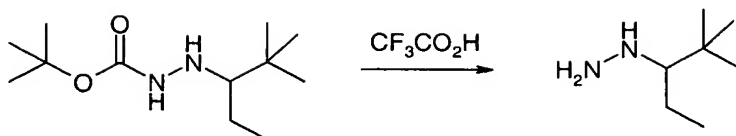
The phases were separated and the organic phase washed once with water, once with 10% NaOH (aq), once with 1 N HCl, and once with saturated aqueous sodium chloride. The solution was dried over magnesium sulfate, filtered, and evaporated to give a white solid. This material was triturated with hexane to give a white solid, 3-(tert-butyl-dimethyl-silanyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid N'-tert-butyl-hydrazide, (0.50 g, 1.22 mmol) in 69% yield. ¹H-NMR (300 MHz, CD₃COCD₃) δ (ppm): 0.12 (s, 6H), 0.92 (s, 9H), 1.53 (s, 9H), 2.32 (s, 3H), 3.97 (m, 2H), 4.14 (m, 1H), 4.28 (m, 1H), 4.43 (m, 1H), 6.79 (d, 1H), 7.22 (d, 1H); TLC R_f = 0.43 (1:1 ethyl acetate : hexane).



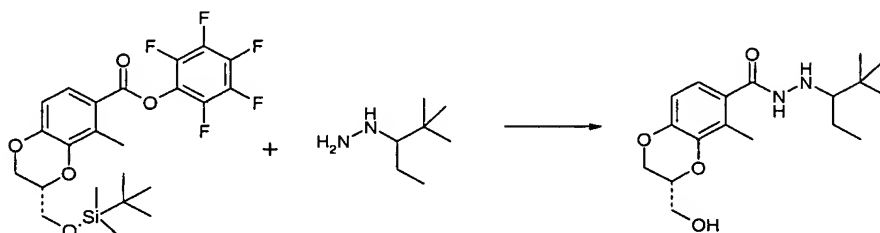
[00292] Tert-butyl,ethylketone (5 g) was mixed with tert-butylcarbazate (6.4 g, 1.1 eq.) in 30 mL of methanol with 3 drops of acetic acid at room temperature. A solid formed within one hour; more methanol was added, and the mixture was stirred at room temperature for three days to produce N'-(1-ethyl-2,2-dimethyl-propylidene)-hydrazinecarboxylic acid tert-butyl ester. ¹H NMR (CDCl₃, 500 MHz) δ (ppm): 7.5 (br, 1H), 2.23 (q, 2H), 1.51 (s, 9H), 1.15 (s, 9H), 1.1 (t, 3H).



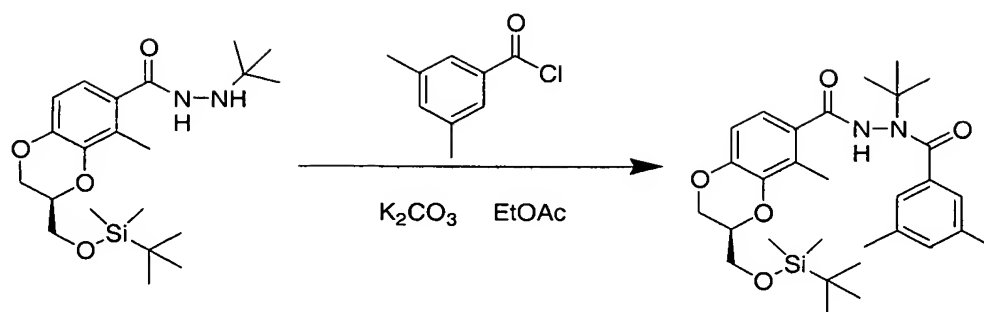
[00293] Approximately 10.3 g N'-(1-ethyl-2,2-dimethyl-propylidene)-hydrazinecarboxylic acid tert-butyl ester was mixed with 4.14 g (3.9 mmol) 10% Pd/C in CH₂Cl₂ under an atmosphere of hydrogen in a Parr shaker for four hours. The reaction was monitored by TLC (I₂ indicator), and the crude product was twice chromatographed using 10% ether in hexane to yield N'-(1-ethyl-2,2-dimethyl-propyl)-hydrazinecarboxylic acid tert-butyl ester. ¹H NMR (CDCl₃, 500 MHz) δ (ppm): 6.05 (br, 1H), 3.9 (br, 1H), 2.45 (m, 1H), 1.6 (m, 1H), 1.5 (s, 9H), 1.25 (m, 1H), 1.1 (t, 3H), 0.92 (s, 9H).



[00294] N'-(1-ethyl-2,2-dimethyl-propyl)-hydrazinecarboxylic acid tert-butyl ester (1.31 g, 5.68 mmol) dissolved in ethyl acetate was mixed with 15 mL trifluoroacetic acid in an ice bath, and stirred at room temperature overnight. The reaction mixture was chilled again, and 40 mL cold water was added, resulting in formation of a new oily phase containing (1-ethyl-2,2-dimethyl-propyl)-hydrazine. 10% NaOH (ca. 60 mL) was then added until the mixture remained basic. This ethyl acetate/aqueous mixture was used "as is" in coupling with a pentafluorophenyl ester.

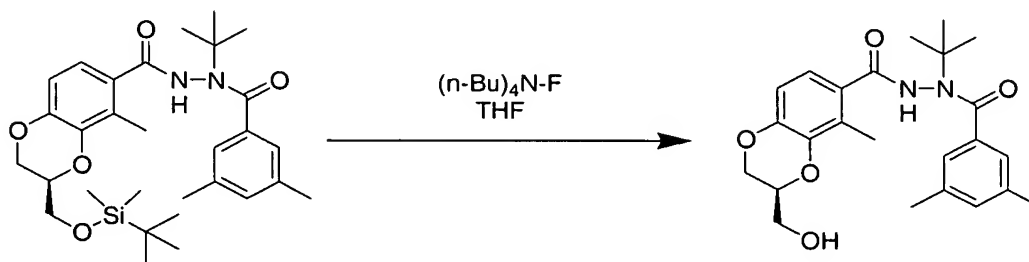


[00295] To the biphasic preparation of (1-ethyl-2,2-dimethyl-propyl)-hydrazine described above, was added a solution of approximately one gram 3(S)-(tert-butyl-dimethyl-silanyloxymethyl)-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carboxylic acid pentafluorophenyl ester dissolved in 10 mL ethyl acetate. The mixture was stirred at room temperature for 2 hours, after which time the aqueous phase was withdrawn and replaced with 1 M K_2CO_3 . The mixture was stirred overnight. The aqueous layer was separated off, and the organic layer was extracted once with 1 M K_2CO_3 and then once with water. The organic phase was dried. The crude product was chromatographed on silica gel using a step gradient of 10% ether in hexane followed by neat ether, yielding 3(S)-hydroxymethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carboxylic acid N'-(1-ethyl-2,2-dimethyl-propyl)-hydrazide. ^1H NMR (500 MHz, CDCl_3) δ (ppm): 7.15 (br, 1H), 6.85 (d, 1H), 6.73 (d, 1H), 4.85 (br, 1H), 4.3 (d, 1H), 4.25 (m, 1H), 4.1 (dd, 1H), 3.87 (m, 2H), 2.41 (m, 1H), 2.29 (s, 3H), 1.8 (br, 1H); 1.65 (m, 1H), 1.3 (m, 1H), 1.45 (t, 3H), 0.98 (s, 9H); TLC: Rf. = 0.23 (1:1 ethyl acetate : hexane).

**RG-115790**

[00296] To a round bottom flask equipped with magnetic stirring was added ethyl acetate (7 mL), 25 wt% aqueous potassium carbonate solution (2.03 g, 3.67 mmol), and 3-(tert-butyldimethylsilyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carboxylic acid N'-tert-butyl hydrazide (0.50 g, 1.22 mmol). To this was added 3,5-dimethylbenzoyl chloride (0.31 g, 1.84 mmol). Stirring was continued for 18 hours. The reaction was diluted with ethyl acetate and the phases separated. The organic phase was washed once with water, once with saturated aqueous sodium chloride, dried over magnesium sulfate, filtered and evaporated. The residue was flash chromatographed on silica gel eluting with 20% ether/hexane, then 50% ether/hexane to give a white solid, 3,5-dimethylbenzoic acid N-tert-butyl-N'-[3-(tert-butyldimethylsilyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carbonyl]-hydrazide, (0.50 g, 0.92 mmol) in 76% yield. ¹H-NMR (300 MHz, CD₃COCD₃) δ (ppm): 0.10 (s, 6H), 0.90 (s, 9H), 1.59 (s, 9H), 1.89 (s, 3H), 2.25 (s, 6H), 3.92 (m, 2H), 4.05 (m, 1H), 4.22 (m, 1H), 4.34 (m, 1H), 6.33 (d, 1H), 6.56 (d, 1H), 7.02 (s, 1H), 7.13 (s, 2H), 9.49 (s, 1H); TLC: R_f = 0.30 (1:1 ether/hexane).

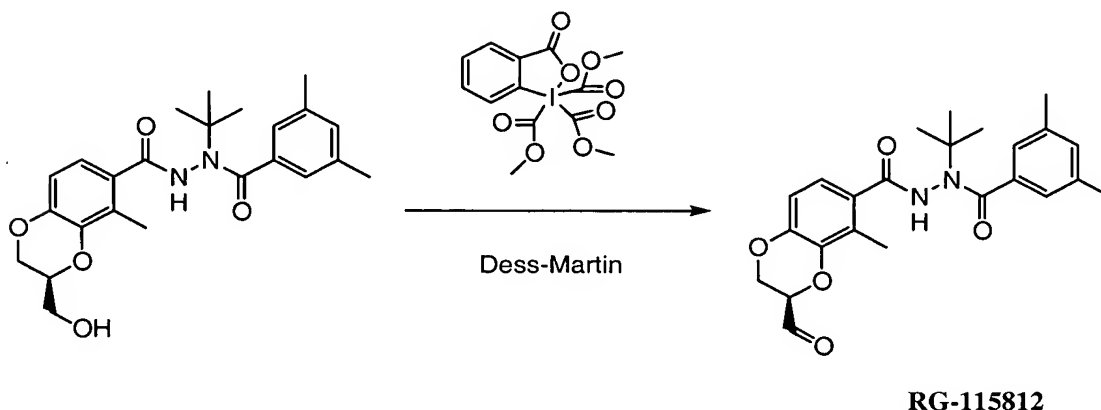
1.8 Preparation of RG-115789

**RG-115789**

[00297] To a round bottom flask equipped with magnetic stirring was added 3,5-dimethylbenzoic acid N-tert-butyl-N'-[3-(tert-butyldimethylsilyloxymethyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carbonyl]-hydrazide (0.26 g, 0.48 mmol), THF (5.0 mL), and 1 M solution of tetrabutylammonium fluoride (TBAF) in THF (1.39 mL, 1.39 mmol). The mixture was stirred at room temperature for

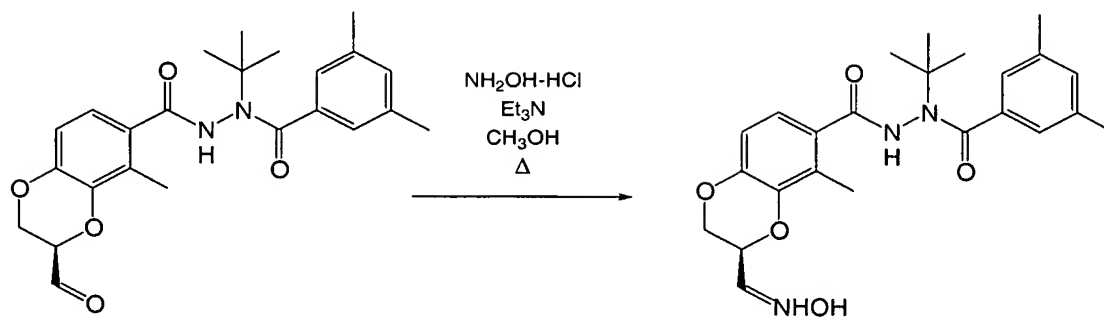
three hours. The THF was evaporated and the residue was taken up in ethyl acetate. This solution was washed once with water, once with saturated aqueous sodium chloride solution, dried over magnesium sulfate, filtered, and evaporated. The residue was triturated with hexane to give a white solid, 3,5-dimethylbenzoic acid N-tert-butyl-N'-(3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carbonyl)-hydrazide, (0.205 g, 0.48 mmol) in 100% yield. $^1\text{H-NMR}$ (300 MHz, CD_3COCD_3) δ (ppm): 1.59 (s, 9H), 1.89 (d, 3H), 2.25 (s, 6H), 3.79 (m, 2H), 4.04 (m, 1H), 4.18 (m, 1H), 4.33 (m, 1H), 6.33 (dd, 1H), 6.56 (dd, 1H), 7.03 (s, 1H), 7.12 (s, 2H), 9.51 (s, 1H); TLC R_f = 0.03 (ether/hexane).

1.9 Preparation of RG-115812



[00298] In a 50 ml round bottom flask, 1.00 g (2.35mm) of 3,5-dimethylbenzoic acid N-tert-butyl-N'-(3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carbonyl)-hydrazide was dissolved in 10 mL of CH_2Cl_2 and 8.0 g of commercially available Dess-Martin reagent were added. The reaction was stirred at room temperature for 24 hours. The reaction mixture was transferred to a separatory funnel with CH_2Cl_2 and extracted with dilute aqueous NaHCO_3 , then with dilute sodium thiosulfate ($\text{Na}_2 \text{S}_2\text{O}_3$) to quench the oxidizing agent. The organic phrase was dried with M_gSO_4 and evaporated to dryness, to yield 1.32 g of product. $^1\text{H NMR}$ (300 MHz, CDCl_3) δ (ppm): 9.70 (s,1H), 7.50(s,1H), 7.05 (s,2H), 6.95 (s, 1H), 6.6 (m, 1H) 6.1-6.2 (q, 1H), 4.6 (d, 2H), 4-4.3 (m, 3H), 2.27 (s, 6H), 2.02 – 2.06 (d, 3H), 1.60 (s, 9H); TLC R_f = 0.13 (1:1 ethyl acatate:hexane) .

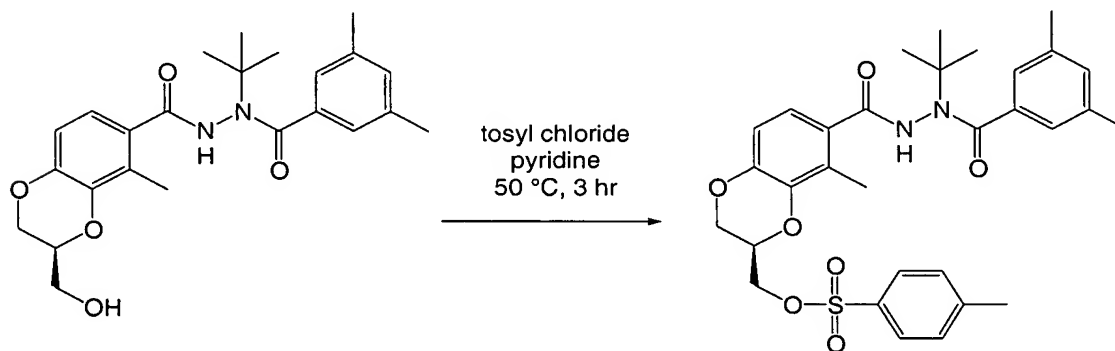
1.10 Preparation of RG-115814



RG-115814

[00299] To a 100 mL round bottom flask, containing 100mg (0.24mm) of 3,5-dimethylbenzoic acid N-tert-butyl-N'-(3-formyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carbonyl)-hydrazide, added 0.50 g of triethylamine, 70 mg (1 mm) of hydroxylamine hydrochloride and 25 mL of methanol. The reaction mixture was refluxed for 1 hour. The methanol was removed on a rotary evaporator and the residue was dissolved with chloroform and dilute HCl and transferred to a separatory funnel. The chloroform extract was dried and evaporated to dryness. The residue was chromatographed on silica and the product eluted with 40% ethyl acetate in hexane. Evaporation of solvent yielded 85 mg of 3,5-dimethylbenzoic acid N-tert-butyl-N'-(3-(hydroxylimino-methyl)-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carbonyl)-hydrazide. ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.7 (s,1H), 7.4 (d,1H), 7.05 (s,2H), 6.95 (s,1H), 6.55 (d,1H), 6.15 (m, 1H), 4.8 (m, 1H), 4.0-4.4 (m,4H), 2.248 (s,6H), 1.915 (s,3H), 1.574 (s, 9H); TLC: R_f = 0.40 (1:1 ethyl acetate : hexane).

1.11 Preparation of RG-115813

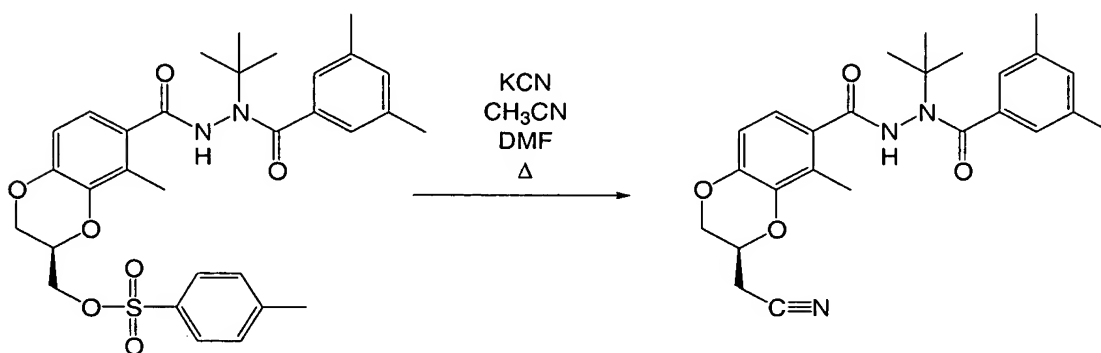


RG-115813

[00300] Into a 100 mL round bottom flask, added 1.00 g (2.35 mm) of 3,5-dimethylbenzoic acid N-tert-butyl-N'-(3-hydroxymethyl-5-methyl-2,3-dihydrobenzo[1,4]dioxine-6-carbonyl)- hydrazide, 538 mg (2.82 mm) of tosyl chloride, and 10 mL of pyridine. The reaction mixture was heated and stirred

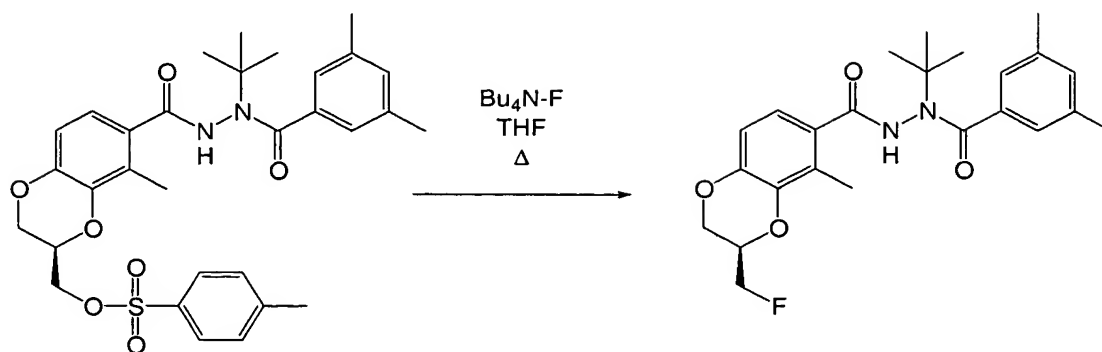
in a 50-60 °C water bath for 3 hours, then stirred at room temperature for 24 hours. The reaction mixture was dissolved in CH₂Cl₂ and first extracted with dilute K₂CO₃, then with dilute HCl (to remove pyridine). The CH₂Cl₂ extract was dried and concentrated to yield about 1.4 g of material. TLC (1:1 ethyl acetate : hexane) gave an R_f of 0.35 for the major compound. The product was purified by column chromatography on silica, eluting with 45% ethyl acetate in hexane to give about 1.1 g of pure toluene-4-sulfonic acid 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxin-2-ylmethyl ester. ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.8 ppm (d, 1H), 7.55 (br s, 1H), 7.36 (d, 1H), 7.05 (s, 2H), 6.95 (s, 1H), 6.5 (m, 1H), 6.1 (q, 1H), 4.0 – 4.4 (m, 5H), 2.45 (s, 3H), 2.25 (s, 6H), 1.85 (d, 3H), 1.58 (s, 9H).

1.12 Preparation of RG-115816

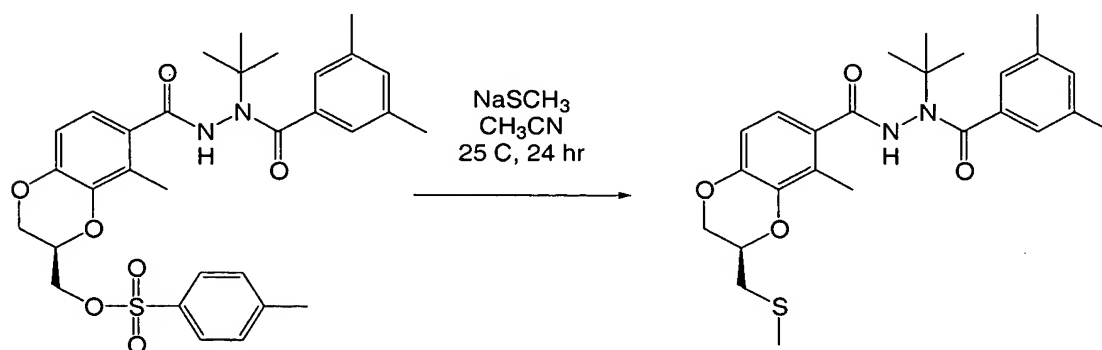


RG-115816

[00301] 300 mg (0.52 mm) of toluene-4-sulfonic acid 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxin-2-ylmethyl ester, 100 mg of KCN, 10 mg of KI, 12 mL of CH₃CN, and 4 mL of DMF were refluxed for 5 hours. The reaction product was concentrated on a rotary evaporator. The reaction product was transferred with ethyl ether and water to a separatory funnel and twice extracted with ether. The ether extract was extracted with water, dried and concentrated to give 0.17 g of a white solid. TLC (1:1 ethyl acetate : hexane) indicated that the product nitrile had a R_f of 0.27. 3,5-Dimethyl-benzoic acid N-tert-butyl-N'-(3-cyanomethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide was purified by silica gel column chromatography, eluting with 45% ethyl acetate in hexane. ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.70 (s, 1H), 7.05 (s, 2H), 6.95 (s, 1H), 6.60 (d, 1H), 6.1-6.2 (m, 1H), 4.0-4.4 (m, 3H), 2.78 (d, 2H), 2.25 (s, 6H), 1.93 (d, 3H), 1.57 (s, 9H).

1.13 Preparation of RG-115815**RG-115815**

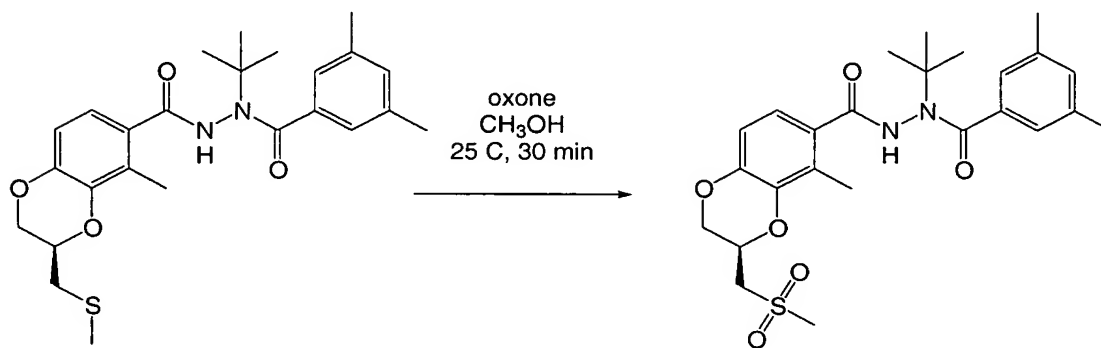
[00302] 250 mg (0.43 mm) of toluene-4-sulfonic acid 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxin-2-ylmethyl ester, 2 mL of 1 M solution of tetrabutylammonium fluoride (TBAF) in THF and 15 mL of THF were refluxed for 4 hours. The reaction mixture was concentrated and re-dissolved with CH_2Cl_2 . The CH_2Cl_2 extract was washed with dilute sodium bicarbonate, dried and concentrated to yield 0.37 g of product. TLC in 1:1 ethyl acetate : hexane gave a Rf. of 0.48 for the major component. Purification by column chromatography on silica gel, eluting with 26% ethyl acetate in hexane yielded 3,5-dimethyl-benzoic acid N-tert-butyl-N'-(3-fluoromethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide (about 0.25 g). ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.80 ppm (s, 1H), 7.02 (s, 2H), 6.95 (s, 1H), 6.5 (d, 1H), 6.1 (t, 1H), 4.70 (d, 1H), 4.5 (d, 1H), 4-4.3 (m, 3H), 2.24 (s, 6H), 1.93 (d, 3H), 1.56 (s, 9H).

1.14 Preparation of RG-115817**RG-115817**

[00303] 400 mg (0.69 mm) of toluene-4-sulfonic acid 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxin-2-ylmethyl ester, 147 mg (2.1 mm) of CH_3SNa , and 15 mL of CH_3CN were stirred at room temperature for 24 hours. The reaction mixture

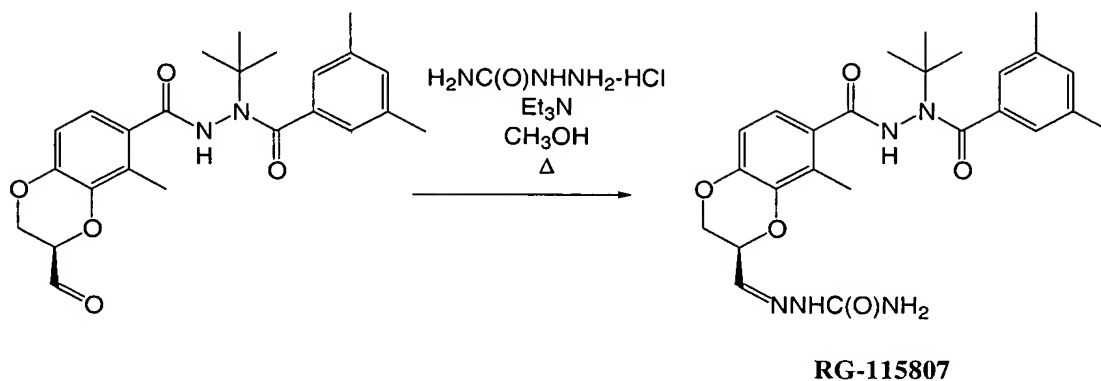
was concentrated to dryness and re-dissolved with CH_2Cl_2 . The CH_2Cl_2 extract was washed with dilute aqueous NaHCO_3 , dried, and concentrated to give 0.26 g of product. TLC indicated a mixture, with a major compound at $R_f = 0.52$ (1:1 ethyl acetate : hexane). Purification by column chromatography on silica gave the desired product, 3,5-dimethyl-benzoic acid N-tert-butyl-N'-(5-methyl-3-methylsulfanylmethyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide, eluted with 24% ethyl acetate in hexane (0.18 g). ^1H NMR: (300 MHz, CDCl_3) δ (ppm): 7.40 (s, 1H), 7.05 (s, 2H), 6.96 (s, 1H), 6.6(d, 1H), 6.1(q, 1H), 4.0-4.3(m, 3H), 2.7-2.9 (m, 2H), 2.27 (s, 6H), 2.21(s, 3H), 1.97 (d, 3H) , 1.59 (s, 9H).

1.15 Preparation of RG-115818

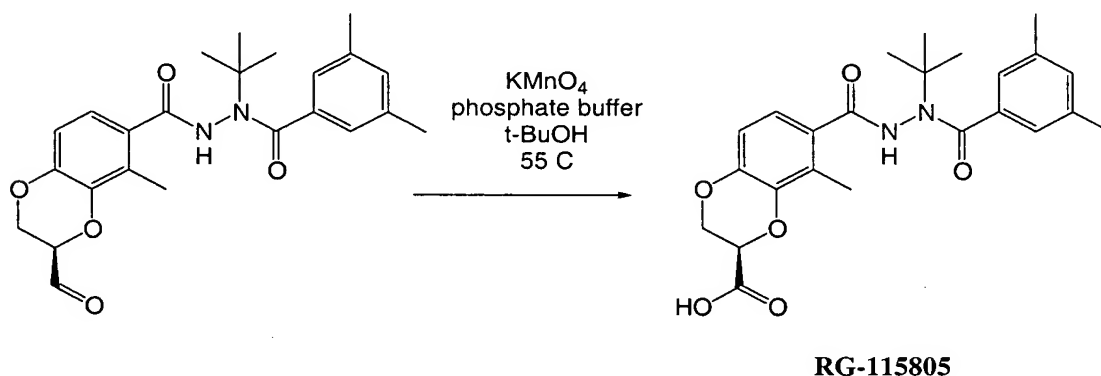


RG-115818

[00304] 125 mg (0.27 mm) of 3,5-dimethyl-benzoic acid N-tert-butyl-N'-(5-methyl-3-methylsulfanylmethyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide were dissolved in 6 mL of CH_3OH . With stirring, 250 mg (0.4 mm) of oxone in 6 mL of water were added, letting stir subsequently for 30 minutes. Methanol was removed on a rotary evaporator, the residue was redissolved in CHCl_3 and water. The aqueous phase was extracted with chloroform, which was then dried and concentrated to yield 130 mg of a white solid. TLC (1:1 ethyl acetate : hexane) showed an R_f of 0.12 for the major product, 3,5-Dimethyl-benzoic acid N-tert-butyl-N'-(3-methanesulfonylmethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide, which was purified by column chromatography on silica gel, eluting with 80% ethyl acetate in hexane. ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.55 (d, 1H), 7.02 (s, 2H), 6.98 (s, 1H), 6.6 (d, 1H), 6.0-6.1 (m, 1H), 4.0-4.5 (m, 3H), 2.8 (d, 2H), 2.25 (s, 6H), 1.92-1.94 (d, 3H), 1.57 (s, 9H).

1.16 Preparation of RG-115807

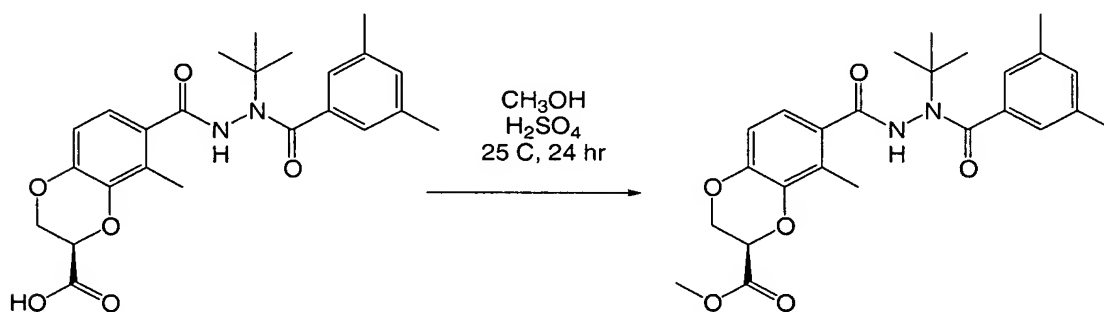
[00305] 200 mg (0.47 mm) of 3,5-dimethyl-benzoic acid N-tert-butyl-N'-(3-formyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide, 0.5g of triethylamine, 210 mg (1.9 mm) of semicarbazide hydrochloride, 10 mL methanol and 2 drops of glacial acetic were added to a 100 mL round bottom flask and refluxed for 4 hours. The reaction mixture was concentrated on the rotary evaporator and redissolved in CHCl_3 . The resultant CHCl_3 solution was extracted twice with dilute NaHCO_3 , dried, and evaporated to yield 0.16 g of crude product. TLC (1:1 ethyl acetate : hexane) showed that major component was at the origin. The product, 3,5-dimethyl-benzoic acid N-tert-butyl-N'-(3-fomyl semicarbazide-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide, was purified by column chromatography on silica gel, eluting with 10% CH_3OH in ethyl acetate. ^1H NMR: (300 MHz, CDCl_3) δ (ppm): $\text{CD}_3\text{OD} - \text{CDCl}_3$: 7.2 (d, 1H), 7.07 (s, 2H), 6.97 (s, 1H), 6.5-6.6 (2d, 1H), 6.2-6.3 (q, 1H), 4.8 (m, 1H) 4.1-4.3 (m, 2H), 2,28 (s, 6H), 1.88 (m, 3H), 1.586 (s, 9H).

1.17 Preparation of RG-115805

[00306] 0.85 g (2 mm) of 3,5-dimethyl-benzoic acid N-tert-butyl-N'-(3-formyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide, 10 mL of warm t-butyl alcohol, and 20 mL of Aldrich phosphate buffer, pH 7.2, #31925-2 were added to a 100 mL round bottom flask, which was placed into a 55 °C water bath. While stirring the reaction mixture, 380 mg (2.4 mm) of potassium

permanganate were slowly added and the reaction was stirred at 55 °C for 7 hours and at room temperature for 24 hours. 10% aqueous NaOH was added until pH was 12 and the black precipitate (MnO_2) was filtered off. The aqueous solution was extracted with ethyl acetate, thereby removing a brown color, transferred to a separatory funnel, acidified with 1N HCL (whereupon product precipitated), and then extracted twice with CHCl_3 . The CHCl_3 extract was dried, evaporated to dryness, and dried in a vacuum oven to yield 0.76 g of product, 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxine-2-carboxylic acid. ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.06 (s, 2H), 6.97 (s, 1H), 6.6 (q, 1H), 6.3 and 6.0 (d+d, 1H), 4.91 (m, 1H), 4.5-4.3 (m, 2H) 2.265 (s, 6H), 2.13 and 1.90 (s+s, 3H) 1.588 (s, 9H); TLC: R_f = 0-0.17, streak (1:1 ethyl acetate : hexane).

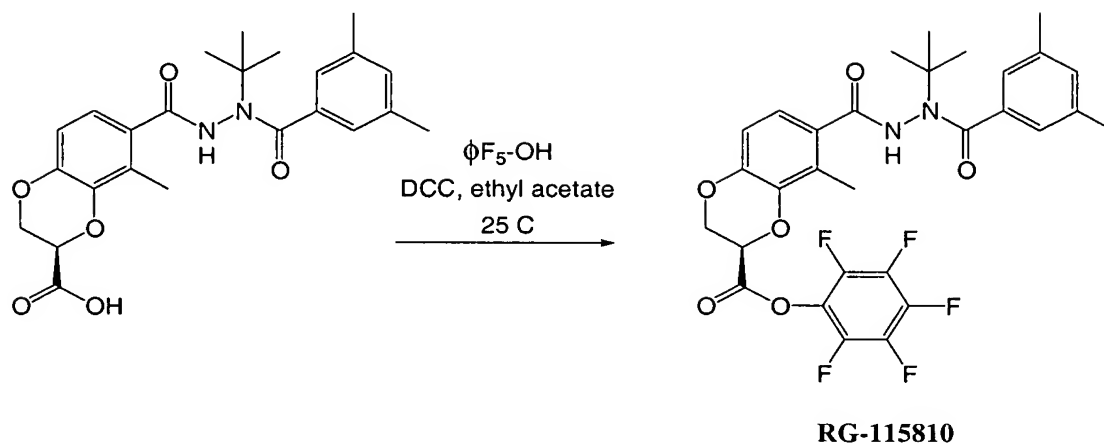
1.18 Preparation of RG-115806



RG-115806

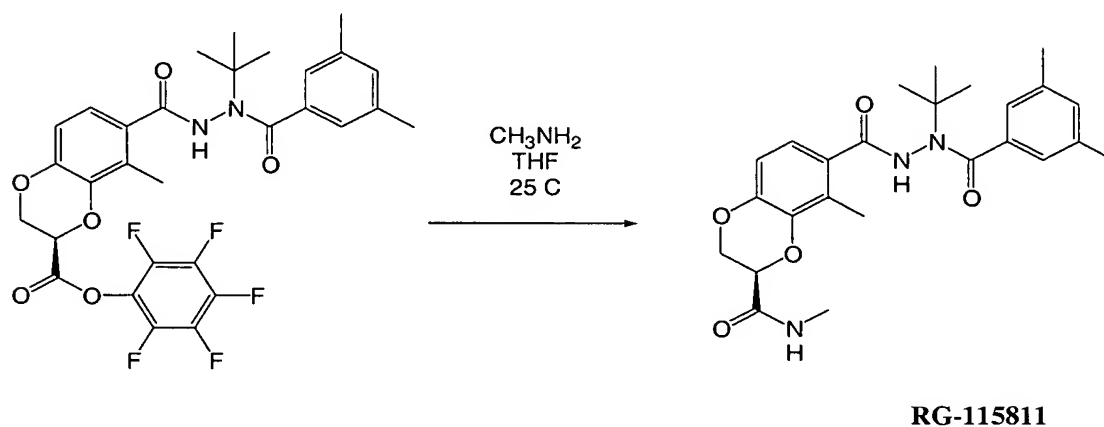
[00307] 190 mg (4.3 mm) of 7-[N'-tert-Butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxine-2-carboxylic acid, 15 mL of CH_3OH and 1 drop of concentrated sulfuric acid were stirred at room temperature for 24 hours. CH_3OH was removed on a rotary evaporator and the residue was dissolved in CHCl_3 and extracted with dilute NaHCO_3 solution. CHCl_3 extract was dried and concentrated to give 0.10 g of white solid. ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.52 (s, 1H), 7.07 (s, 2H), 6.97 (s, 1H), 6.5-6.6 (q, 1H), 6.3 and 6.0 (d+d, 1H), 4.8 (m, 1H), 4.4-4.2 (m, 2H), 3.772 (3H), 2.26 (s, 6H), 2.035 and 1.919 (s+s, 3H), 1.584 (s, 9H). TLC: R_f = 0.35 (1:1 ethyl acetate : hexane).

1.19 Preparation of RG-115810



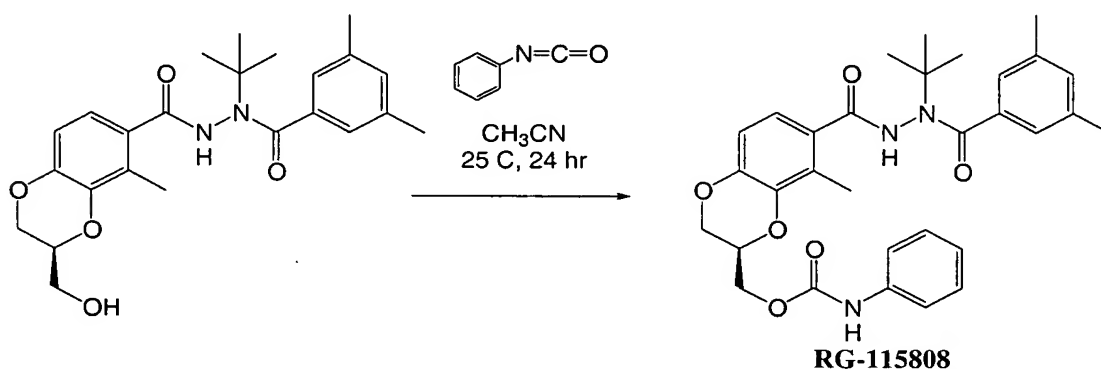
[00308] In a vial, 97 mg (0.32 mm) of 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxine-2-carboxylic acid, 60 mg (0.35 mm) of pentafluorophenol, 60 mg (0.32 mm) of DCC and 3 mL of ethyl acetate were stirred at room temperature for 24 hours. The reaction mixture was transferred to a round bottom flask and evaporated to dryness. The residue was redissolved in CH_2Cl_2 and chromatographed on silica gel. Elution with 40% ethyl acetate in hexane yielded the pentafluorophenol ester, 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxine-2-carboxylic acid pentafluorophenyl ester, in a quantity of 80 mg. ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.60 (d, 1H), 7.03 (s, 2H), 6.95 (d, 1H), 6.6 (q, 1H), 6.3 and 6.0 (d+d, 1H), 4.6 (m, 1H), 4.4 (m, 1H), 2.251 and 2.230 (s+s, 6H), 2.052 and 1.903 (s+s, 3H), 1.58 and 1.574 (s+s, 9H) (many signals split); TLC: R_f = 0.52 (1:1 hexane : ethyl acetate).

1.20 Preparation of RG-115811

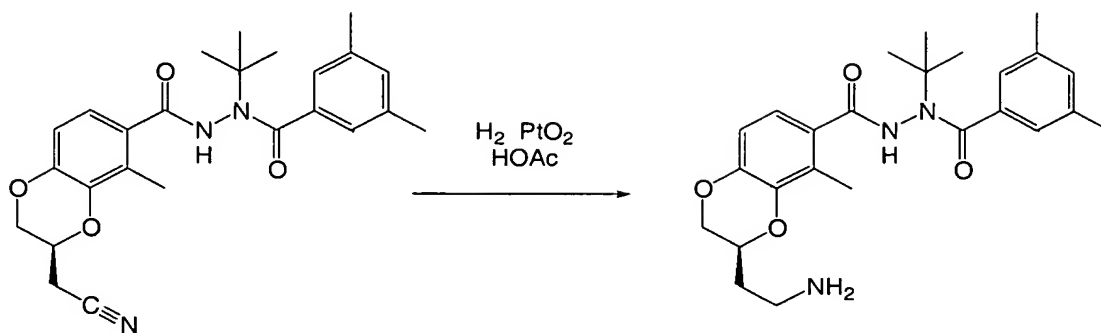


[00309] In a vial, 80 mg (0.2 mm) of 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxine-2-carboxylic acid pentafluorophenyl ester and 2 mL of a 2 M CH_3NH_2 solution in THF were stirred at room temperature for 24 hours. The reaction mixture was transferred to a silica chromatography column. Impurities were eluted with 45% ethyl acetate in hexane and product, 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxine-2-carboxylic acid methylamide, was eluted with 90% ethyl acetate in hexane. ^1H NMR (300 MHz, CDCl_3) δ (ppm): 8.1 (d, 1H), 7.05 (s, 2H), 6.97 (d, 1H), 6.40 (d, 1H), 6.00 (d, 1H), 4.45-4.25 (m, 2H), 3.9-4.0 (m, 1H), 2.88 and 2.86 (s+s, 3H), 2.24 and 2.22 (s+s, 6H), 2.07 and 1.87 (s+s, 3H), 1.59 and 1.57 (s+s, 9H) (several split signals); TLC: R_f = 0.06 (1:1 ethyl acetate : hexane).

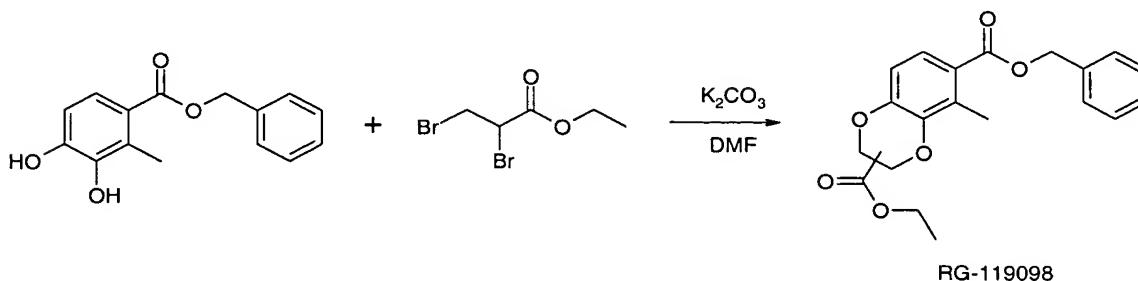
1.21 Preparation of RG-115808



[00310] 120 mg (0.28 mm) of 3,5-Dimethyl-benzoic acid N-tert-butyl-N'-(3-hydroxymethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide, 70 mg (0.55 mm) of phenylisocyanate and 2 mL of CH_3CN were stirred in a vial at room temperature for 24 hours. The CH_3CN was blown off with a stream of N_2 and the residue was triturated with pentane. The supernatant was removed and the residual pentane blown off with N_2 to give 0.15 g of product phenyl-carbamic acid 7-[N'-tert-butyl-N'-(3,5-dimethyl-benzoyl)-hydrazinocarbonyl]-8-methyl-2,3-dihydro-benzo[1,4]dioxin-2-ylmethyl ester. ^1H NMR (300 MHz, CDCl_3) δ (ppm): 7.65 (d, 1H), 7.4-7.2 (m, 5H), 7.1 (d, 1H), 7.06 (s, 2H), 6.98 (s, 1H), 6.86 (br s, 1H), 6.53 (d, 1H), 6.12 (d, 1H), 4.5-4.2 (m, 3H), 4.0 (m, 2H), 2.23 (s, 6H), 1.94 (s, 3H), 1.58 (s, 9H); TLC: R_f 0.4-0.5, streak (1:1 ethyl acetate hexane).

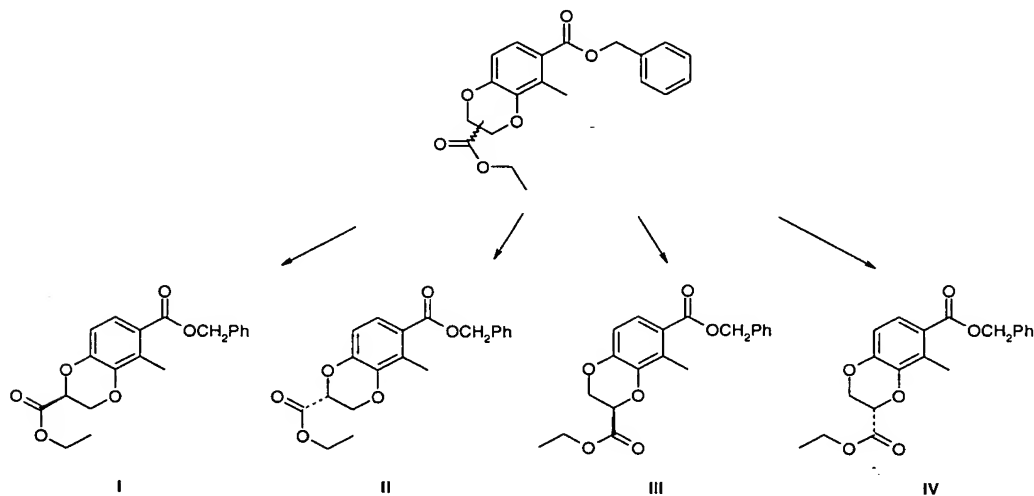
1.22 Preparation of RG-115809**RG-115809**

[00311] 150 mg (0.34 mm) of 3,5-Dimethyl-benzoic acid N-tert-butyl-N'-(3-cyanomethyl-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl)-hydrazide were dissolved in 8 mL of glacial acetic acid and added to a Parr hydrogenation bottle, together with about 16 mg of PtO₂. Parr hydrogenation was conducted for 4 hours and the reaction mixture was filtered. The contents were transferred to a round bottom flask with CHCl₃ and ethyl acetate, and reaction mixture was concentrated to dryness. The residue was redissolved with CHCl₃ and 0.1N KOH, and transferred to a separatory funnel. The aqueous phase was again extracted with CHCl₃ and the CHCl₃ extract was dried, and concentrated to yield 0.14 g of product, 3,5-dimethyl-benzoic acid N'-[3-(2-amino-ethyl)-5-methyl-2,3-dihydro-benzo[1,4]dioxine-6-carbonyl]-N-tert-butyl-hydrazide. ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.8 (d, 1H), 7.00 (s, 2H), 6.95 (s, 1H), 6.6-6.5 (m, 1H), 6.2-6.1 (m, 1H), 4.4-3.8 (m, 3H), 2.9-2.75 (m, 2H), 2.244 (s, 6H), 1.92 (t, 3H), 1.7 (m, 2H), 1.565 (s, 9H); TLC: R_f = 0.24 (1:1 ethyl acetate : hexane).

1.23 Preparation of RG-119098**RG-119098**

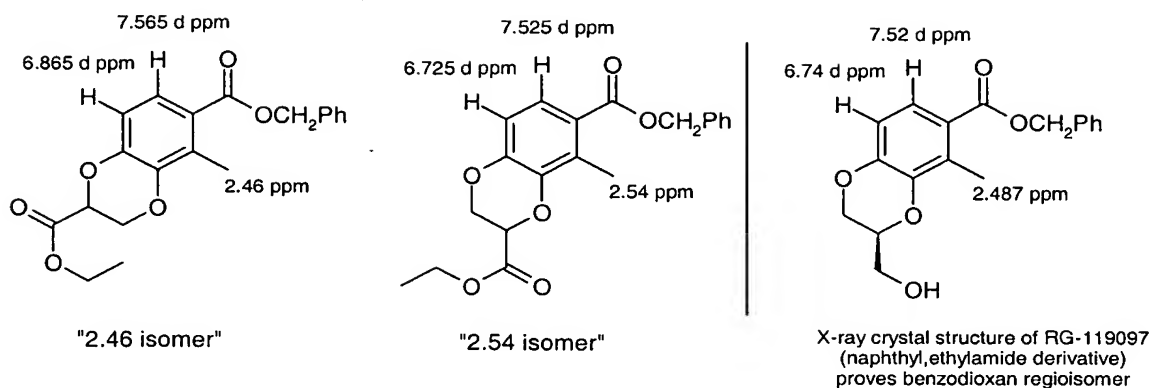
[00312] 16.7 g (64.7 mmol) of benzyl 2-methyl-3,4-dihydroxybenzoate were mixed with ethyl 2,3-dibromopropionate (20.17 g, 77.6 mmol), potassium carbonate (10.72 g, 77.6 mmol) and DMF (216 mL) and heated to 40-45 °C for 4 hours. The reaction mixture was diluted with ether, washed once with water and thrice with brine. The organic layer was dried over magnesium sulfate and evaporated. A small sample was purified by flash chromatography, eluting with 10% ether/methylene

chloride. Kugelrohr distillation under high vacuum, and heating up to 200 °C does not volatilize the desired product, which remains in the distillation flask as an orange oil. Nonetheless, such treatment provided some purification of 8-methyl-2,3-dihydro-benzo[1,4]dioxine-2 (or 3),7-dicarboxylic acid 7-benzyl ester 2 (or 3)-ethyl ester. ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.565 and 7.525 (d+d, 1H), 7.4 (m, 5H), 6.865 and 6.725 (d+d 1H), 5.3 (s, 2H), 4.9 (m, 1H), 4.4 (m, 2H), 4.3 (m, 2H), 2.54 and 2.46 (s+s, 3H), 1.3 (m, 3H).



[00313] The approximately 1:1:1:1 mixture of benzodioxan regio- and stereoisomers which RG-119098 comprises, 8-methyl-2,3-dihydro-benzo[1,4]dioxine-2 (or 3),7-dicarboxylic acid 7-benzyl ester 2 (or 3)-ethyl ester, were resolved from one another to baseline resolution and purified on a multi-gram scale using a Chiralcel® OD-H® HPLC chiral chromatography column, serial # ODH0CE-CB035. The mobile phase was 97.5:2.5 hexane:ethanol at a flow rate of 1 ml/min. at 25 °C. Each of the four isomers was isolated from the remaining three. The work was performed by Chiral Technologies, 730 Spring Drive, Exton PA.

[00314] In another experiment, approximately 12 g of benzyl ester RG-119098 was purified on silica gel. Elution with hexane yielded approximately 10 g of a roughly 1:1 mixture of 2 regioisomers. These were designated the “2.46 isomer” and the “2.54 isomer”, based upon the absorbance of the benzylic CH₃ group. Continued elution with 10% ethyl acetate in hexane produced a sample that was enriched in the 2.46 isomer in a ratio of approximately 2:1 relative to the 2.54 isomer. Re-chromatography of this material in a gradient of 6-9% ethyl acetate in hexane demonstrated further enrichment of the 2.46 isomer. Thus, the fraction eluted with 8% ethyl acetate in hexane was comprised of a 6:1 ratio and the 9% fraction was comprised of an 8:1 ratio of the 2.46:2.54 isomers.



[00315] The isomers are tentatively assigned as in the diagram based on analogous ^1H NMR (300 MHz, CDCl_3) signals to those of a related 2-hydroxymethylbenzodioxan of known regiochemistry, based on correlation to the crystal structure of RG-119097. Comparison of the ^1H NMR of isolated isomers comprising RG-119098 to the corresponding ethyl ester correlated to RG-119097 (isomer III indicated in diagram) would provide an unambiguous regiochemical assignment of each isomer.

EXAMPLE 2: BIOLOGICAL TESTING OF COMPOUNDS

[00316] The ligands of the present invention are useful in various applications including gene therapy, expression of proteins of interest in host cells, production of transgenic organisms, and cell-based assays.

27-63 Assay

Gene Expression Cassette

[00317] GAL4 DBD (1-147)-CfEcR(DEF)/VP16AD- β RXREF-LmUSPEF: The wild-type D, E, and F domains from spruce budworm *Choristoneura fumiferana* EcR ("CfEcR-DEF"; SEQ ID NO: 1) were fused to a GAL4 DNA binding domain ("Gal4DBD1-147"; SEQ ID NO: 2) and placed under the control of a phosphoglycerate kinase promoter ("PGK"; SEQ ID NO: 3). Helices 1 through 8 of the EF domains from *Homo sapiens* RXR β ("HsRXR β -EF"; nucleotides 1-465 of SEQ ID NO: 4) and helices 9 through 12 of the EF domains of *Locusta migratoria* Ultraspiracle Protein ("LmUSP-EF"; nucleotides 403-630 of SEQ ID NO: 5) were fused to the transactivation domain from VP16 ("VP16AD"; SEQ ID NO: 6) and placed under the control of an elongation factor-1 α promoter ("EF-1 α "; SEQ ID NO: 7). Five consensus GAL4 response element binding sites ("5XGAL4RE"; comprising 5 copies of a GAL4RE comprising SEQ ID NO: 8) were fused to a synthetic TATA minimal promoter (SEQ ID NO: 9) and placed upstream of the luciferase reporter gene (SEQ ID NO: 10).

[00318] CHO cells were transiently transfected with transcription cassettes for GAL4 DBD (1-147) CjEcR(DEF) and for VP16AD β RXREF-*Lm*USPEF controlled by ubiquitously active cellular promoters (PGK and EF-1 α , respectively) on a single plasmid. Stably transfected cells were selected by Zeocin resistance. Individually isolated CHO cell clones were transiently transfected with a GAL4 RE-luciferase reporter (pFR Luc). 27-63 clone was selected using Hygromycin.

Treatment with Ligand

[00319] Cells were trypsinized and diluted to a concentration of 2.5×10^4 cells mL. 100 μ L of cell suspension was placed in each well of a 96 well plate and incubated at 37°C under 5% CO₂ for 24 h. Ligand stock solutions were prepared in DMSO and diluted 300 fold for all treatments. Dose response testing consisted of 8 concentrations ranging from 33 μ M to 0.01 μ M.

Reporter Gene Assay

[00320] Luciferase reporter gene expression was measured 48 h after cell treatment using Bright-GloTM Luciferase Assay System from Promega (E2650). Luminescence was detected at room temperature using a Dynex MLX microtiter plate luminometer. EC₅₀s were calculated from dose response data using a three-parameter logistic model.

[00321] The results of the assays are shown in Table 1. Each assay was conducted in two separate wells, and the two values were averaged. Relative Max FI was determined as the maximum fold induction of the tested ligand (an embodiment of the invention) observed at any concentration relative to the maximum fold induction of GS-TM-E (3,5-Dimethyl-benzoic acid N-tert-butyl-N'-(2-ethyl-3-methoxy-benzoyl)-hydrazide) observed at any concentration.

Table 1. Biological Assay Results for Compounds

Compound	EC ₅₀ (μ M)/relative Max FI 27-63 Assay
RG-115789	1.85/0.81
RG-115790	> 33/0.0
RG-115805	> 33/0.01
RG-115806	~ 33/0.25
RG-115807	3.34/0.80
RG-115808	1.12/2.37
RG-115809	3.25/1.09
RG-115810	~ 20/1.23
RG-115811	> 33/0.24
RG-115812	2.68/0.86
RG-115813	> 33/0.01
RG-115814	0.32/1.0

Compound	EC ₅₀ (μM)/relative Max FI 27-63 Assay
RG-115815	0.3/0.738
RG-115816	4.22/0.85
RG-115817	2.90/0.95
RG-115818	5.26/0.90
RG-115843	0.85/0.73
RG-115844	1.23/0.352
RG-115853	0.10/0.898
RG-115854	0.20/0.917
RG-115845	4.47/0.52
RG-115855	0.0267/1.16
RG-115860	Avg FI=2307 at 33 μM
RG-115877	2.25/1.48
RG-115878	0.10/1.05
Reference:	
RG-102317	0.102
GS TM -E ligand	0.288

GSTM-E ligand=3,5-Dimethyl-benzoic acid N-tert-butyl-N'-(2-ethyl-3-methoxy-benzoyl)-hydrazide

[00322] In addition, one of ordinary skill in the art is also able to predict that the ligands disclosed herein will also work to modulate gene expression in various cell types described above using gene expression systems based on group H and group B nuclear receptors.